HAL
open science

RNA sequencing and analysis of three Lupinus nodulomes provide new insights into specific host-symbiont relationships with compatible and incompatible Bradyrhizobium strains

Jean Keller, J. Imperial, T. Ruiz-Argüeso, Kaïna Privet, Oscar Lima, Sophie Michon-Coudouel, Marine Biget, Armel Salmon, Abdelkader Aïnouche, Francisco Cabello-Hurtado

To cite this version:

Jean Keller, J. Imperial, T. Ruiz-Argüeso, Kaïna Privet, Oscar Lima, et al.. RNA sequencing and analysis of three Lupinus nodulomes provide new insights into specific host-symbiont relationships with compatible and incompatible Bradyrhizobium strains. Plant Science, 2018, 266, pp.102-116. 10.1016/j.plantsci.2017.10.015 . hal-01671540

HAL Id: hal-01671540

https://univ-rennes.hal.science/hal-01671540v1

Submitted on 16 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ACCEPTED MANUSCRIPT

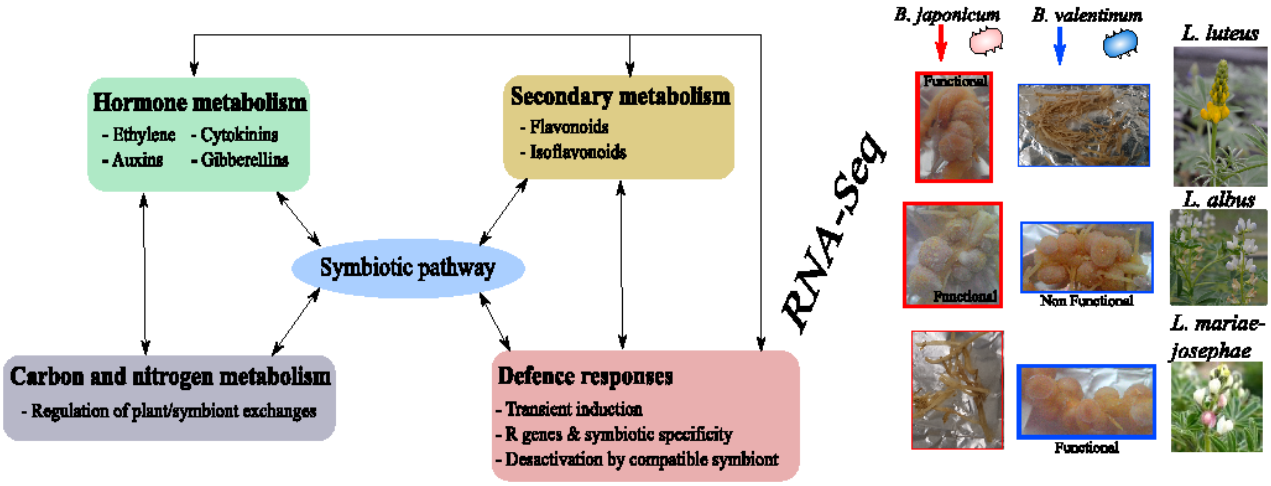
Title: RNA sequencing and analysis of three *Lupinus* nodulomes provide new insights into specific host-symbiont relationships with compatible and incompatible *Bradyrhizobium* strains

J. Keller[1], J. Imperial[2,3], T. Ruiz-Argueso[2], K. Privet[1], O. Lima[1], S. Michon-Coudouel[4], M. Biget[4], A. Salmon[1], A. Aïnouche[1], F. Cabello-Hurtado[1*]

[1]UMR CNRS 6553 Ecobio, OSUR (Observatoire des Sciences de l'Univers de Rennes), Université de Rennes 1, 35042 Rennes, France ; [2]Centro de Biotecnología y Genómica de Plantas (CBGP), Universidad Politécnica de Madrid (UPM) – Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), 28223 Pozuelo de Alarcón, Madrid, Spain; [3]Instituto de Ciencias Agrarias, Consejo Superior de Investigaciones Científicas (CSIC), 28006 Madrid, Spain ; [4] Environmental and Human Genomics Platform, OSUR (Observatoire des Sciences de l'Univers de Rennes), Université de Rennes 1, 35042 Rennes, France

*Corresponding author. Phone: +33 2 23235022. Email: francisco.cabello-hurtado@univ-rennes1.fr

Graphical abstract



Abstract

Nitrogen fixation in the legume root-nodule symbiosis has a critical importance in natural and agricultural ecosystems and depends on the proper choice of the symbiotic partners. However, the genetic determinism of symbiotic specificity remains unclear. To study this process, we inoculated three *Lupinus* species (*L. albus*, *L. luteus*, *L. mariae-josephae*), belonging to the under-investigated tribe of Genistoids, with two *Bradyrhizobium* strains (*B. japonicum*, *B. valentinum*) presenting contrasted degrees of symbiotic specificity depending on the host. We produced the first transcriptomes (RNA-Seq) from lupine nodules in a context of symbiotic specificity. For each lupine species, we compared gene expression between functional and non-functional interactions and determined differentially expressed (DE) genes. This

ACCEPTED MANUSCRIPT

revealed that *L. luteus* and *L. mariae-josephae* (nodulated by only one of the *Bradyrhizobium* strains) specific nodulomes were richest in DE genes than *L. albus* (nodulation with both microsymbionts, but non-functional with *B. valentinum*) and share a higher number of these genes between them than with *L. albus*. In addition, a functional analysis of DE genes highlighted the central role of the genetic pathways controlling infection and nodule organogenesis, hormones, secondary, carbon and nitrogen metabolisms, as well as the implication of plant defence in response to compatible or incompatible *Bradyrhizobium* strains.

Keywords: lupines, symbiotic specificity, RNA-Seq, nodulation phenotypes, functional markers

1. Introduction

Nitrogen is one of the most important macronutrients for plant life and development. Access to assimilable nitrogen is particularly challenging for land plants that are sessile organisms unable to move following the soil nitrogen availability. One of the most complex and efficient strategy developed by land plants to overcome the risk of nitrogen starvation is the establishment of mutualistic nitrogen-fixing symbioses between plants from diverse lineages and different bacteria able to fix atmospheric nitrogen; the latter benefiting in return from plant carbohydrates [1]. Among the land plants, the legumes (Fabaceae; Angiosperms) encompass the majority of species able to form nitrogen-fixing symbioses. With more than 19,000 species spread within approximately 770 genera adapted to a wide range of ecogeographical conditions, Fabaceae is the third largest angiosperm family [2]. The majority of legumes are able to associate with nitrogen-fixing bacteria (members of the α- and β-proteobacteria) which are collectively termed rhizobia [3]. This ability to form nitrogen-fixing symbioses confers the Fabaceae a major role in natural ecosystems and makes them of a great interest for agriculture [4]. In addition, they provide valuable protein sources for feed and food and are currently considered as nutraceutics [4,5].

Nitrogen fixation occurs in specialised root structures called nodules providing a suitable environment with low oxygen tension and high access to plant-derived sources of carbon and energy that are necessary for nitrogen fixation by the bacterial nitrogenase [6]. Accordingly, legume symbiosis has been extensively studied during the last three decades, which significantly increased our understanding of the genetic pathway underlying the establishment and function of root nodule symbiosis (RNS) [7,8]. Symbiosis is the result of a molecular dialogue between plant host and symbiont. Plant hosts release a cocktail of symbiotic effectors (mainly composed of (iso)flavonoids) to the rhizosphere that act as selectors of compatible bacteria [9]. Plant effectors activate a gene cascade in rhizobia resulting in the release of lipochitooligosaccharides called NOD factors [9] that are perceived by specific transmembrane receptors (*NFR1, NFR5* and putatively *SymRK*) at the surface of root epidermal cells [10,11]. The activation of plant receptors induces two downstream programs. The first occurs in the epidermal cells and regulates the infection of root by rhizobia. The second is located in the cortical cells and controls the cell division leading to the nodule organogenesis [7]. NOD factor perception by transmembrane receptors induces the transduction of a signal to the cell nucleus through nucleoporins (*NUP85, NUP133, NENA*) where it

ACCEPTED MANUSCRIPT

activates calcium spiking, Ca^{2+} influx and membrane depolarization involving different cation channels (*MCA8, CASTOR, POLLUX*) [12–14]. Calcium spiking is then decoded by a calcium calmodulin dependent protein (CCAMK) associated with a nuclear coiled-coil protein (CYCLOPS) [15]. This complex regulates transcription factors and downstream genes (*NIN, NSP1, NSP2, ERF, LHK, VAPYRIN*) involved in both epidermal and cortical programs [7]. This genetic pathway is interconnected with several other metabolic pathways, such as hormone biosynthesis and signalling, secondary metabolism, as well as the regulation of carbon and nitrogen fluxes and the plant defence response system [7,8,16]–[21]. Taken together, all the metabolic pathways involved in the root nodule symbiosis and their interactions make this process highly complex and many specific details remain to be discovered. Improving our knowledge of the RNS regulation and legume-symbiont relationships is crucial for a better understanding of natural ecosystems and could be very useful in a context of crop improvement. To date, RNS is extensively studied in a few model plants such as soybean (*Glycine max*), barrel medic (*Medicago truncatula*) or birdsfoot trefoil (*Lotus japonicus*). Although these species represent different nodulation systems (e.g indeterminate *vs* determinate nodules), they all belong to the same legume lineage, whereas very little is known about the symbiosis in other legume lineages. Within the Papillionoid legumes, the *Lupinus* genus, which is member of the Genistoids tribe, is of particular ecological and agronomical interest [17]. This genus encompasses hundreds of species mainly spread in the New World, whereas around 20 species and subspecies are found in the Old World [18]. Recently, attention paid to this genus drastically increased because of its beneficial proprieties for human health and nutrition [5], and the recent release of novel genomic resources and of a first draft genome [19–21]. Lupines are also particularly suitable for the symbiosis studies since they are, to date, the only known legume genus unable to form arbuscular mycorrhizal symbiosis, and since their RNS present singular features, such as the absence of infection thread and specific lupinoid nodules [22]. Lupines are mainly nodulated by diverse *Bradyrhizobium* strains, mainly belonging to *B. japonicum* and *B. canariense* in the Old World, and to *B. japonicum* in the New World [23]. Although most lupine species appear to be nodulated by various strains of *Bradyrhizobium*, some cases of symbiotic specificity have been reported. The best-known example of symbiotic specificity has been observed in a newly described species (*L. mariae-josephae*), endemic from the Valencia region in Spain and adapted to alkaline and calcareous soils [24]. This species is nodulated by *Bradyrhizobium valentinum* [25]. Previous cross-inoculation studies revealed that *L. mariae-josephae* was unable to form functional RNS with different isolates of *B. japonicum* or *B. elkanii*, and that *B. valentinum* is not able to successfully nodulate other closely-related lupine species[25].

Symbiotic specificity is one of the most striking aspects of RNS, and while it has been observed many times among the Fabaceae family, the mechanisms underlying this process are still unclear [25]. This complex process occurs at different stages of the RNS, from the pre-infection molecular exchanges to the nitrogen fixation inside the nodules [26]. To date, different factors controlling host range have been identified, such as the specific mutual recognition of both partners through the plant (flavonoids) and rhizobia (NOD factors, surface exopolysaccharides…) specific effectors [26,27], and the homo- or heterozygote status of genes encoding bacterial effector receptors [28,29]. Symbiotic specificity has also been observed in latest stages of RNS, where rhizobia strains successfully infected the root and induced

ACCEPTED MANUSCRIPT

nodule organogenesis, but were unable to fix nitrogen in some hosts, whereas these strains induced a fully functional RNS with other hosts [30–32]. Thus, symbiotic specificity may operate at different stages of the RNS, and may involve different genes from various functional categories. In addition, a better understanding of this process is crucial and could open new ways to improve the nitrogen fixation of legume crops by modifying their symbiont range.

In this context, we assembled and analysed transcriptomes from roots and nodules of three closely-related Old World lupine species (*L. albus*, *L. luteus* and *L. mariae-josephae*) inoculated with compatible, incompatible or partially compatible *Bradyrhizobium* strains. Transcriptome analysis allowed the study of the RNS genetic pathway and its expression pattern in the under investigated genus *Lupinus* which is of growing interest. These analyses revealed that, although lupines seem to have specific nodulation features, canonical symbiotic genes known from other model legumes are expressed during symbiosis. In addition, the analyses of differentially expressed genes provide new insights into the global response of lupines to the compatible or incompatible bacterial species. They also highlighted that, along with symbiosis-related genes, hormone and defence-related genes, carbon, nitrogen as well as secondary metabolisms are deeply affected by symbiont compatibility.

2. Materials & Methods

2.1. Experimental design

A cross-inoculation experiment between three closely related Euro-Mediterranean *Lupinus* species and two compatible or incompatible *Bradyrhizobium* strains was conducted. The three lupine species were represented by the following accessions: M20/USDA PI250094 (from Egypt) for *L. albus*, LL049/INRA-Dijon-FR (Pisa, Italy) for *L. luteus*, and Lmj/CIEF/V39Q (Llombai, Valencia, Spain) for *L. mariae-josephae*. Following the results from previous investigation on the compatibility between Mediterranean lupines and various *Bradyrhizobium* strains [25], two bacterial strains were used in this experiment: *B. japonicum* isolate ISLU101, which efficiently nodulates *L. luteus* and *L. albus* (also compatible with most Mediterranean lupines) but not *L. mariae-josephae*; and *B. valentinum* isolate LmjC, which only establishes fully functional root nodule symbiosis with *L. mariae-josephae*, and that results in the formation in *L. albus* of white nodules presenting a very low level of nitrogen fixation (Supplementary Figure S1; [25]). The cross-inoculation experiment was carried out as described in [25] at CBGP (UPM-INIA). Briefly, sterile Leonard jar units containing N-free Jensen's solution were used [33]. Each unit was filled with autoclaved vermiculite, and two plants were grown in each from axenically-germinated seedlings. Plants were grown in the Leonard jar units under bacteriologically controlled conditions in a greenhouse under natural light supplemented with artificial light (16/8h day/night, at 25/23°C). After four weeks of growth, roots and nodules were harvested (Supplementary Figure S1) and immediately frozen in liquid nitrogen. For each lupine/*Bradyrhizobium* pair, three replicates were separately collected.

2.2. RNA extraction and sequencing

RNA was extracted from frozen nodules and their surrounding root tissue. Here, all nodules were considered regardless of the development stage. Where no nodules were formed (such as in the *L.*

ACCEPTED MANUSCRIPT

luteus/LmjC and *L. mariae-josephae*/ISLU101 cases), RNA was extracted from roots. RNAs from the 18 samples (3 lupine species*2conditions*3replicates) were extracted from 150 mg of frozen material using the RNeasy Plant Mini Kit from Qiagen following the manufacturer's recommendations. Potential DNA contamination was removed by a DNAse treatment using the DNAse Ambion Kit (Life Technologies, United States) and following the manufacturer's recommendations. Quantity and quality of RNA samples were then checked with a Nanodrop ND1000 spectrophotometer (Thermo Fisher Scientific, United States) and an Agilent BioAnalyzer (Agilent Technologies, United States), respectively. Finally, all samples were subjected to high-throughput sequencing on the Illumina HiSeq 2500 platform at BGI (Hong-Kong) with a library size of 500 bp. In order to preferentially sequence the plant RNA, RNA-Seq has been performed after poly(A+) enrichment. Bacterial remaining sequences were then identified by mapping reads against *B. valentinum* LmjM3 and *B. japonicum* genomes (assemblies accession ASM144040v1 and ASM28437v1) using Bowtie2 [34] and not further considered.

2.3. Transcriptome assembly and annotation

Quality of paired-end (PE) reads was first checked using FASTQC software (v0.11.4) [35]. For each species, reads from the three replicates of the two conditions were pooled prior to the assembly. For each *Lupinus* species, transcriptome assembly was achieved using Trinity v2.0.6 [36] with the default parameters on a high-performance computing cluster (Genouest Bioinformatics, Rennes, France). Then, potential chimeras were removed using a Perl script from the GenoToul Bioinformatic platform. For each species, reads of each replicate were individually mapped back to the transcriptome using Bowtie 2 v2.2.8 [34] in end-to-end mode with no gaps and the no-mixed and no-discordant options, as well as the –k option fixed to 200 to allow multiple reads mapping. Then contigs were clustered using Corset software v1.04 [37]. This allows the hierarchical clustering of contigs based on their number of shared reads and their expression ratio between the two conditions. For each cluster, the longest contig was selected as representative and used for further analysis. The clustered transcriptomes were subjected to an analysis of completeness using Benchmarking Universal Single-Copy Orthologs (BUSCO [38]) and the plants orthologs database (odb9, available at http://busco.ezlab.org/). Additionally, Corset produces a matrix containing the number of reads for each cluster that was used for downstream expression analysis [37].

Each transcriptome was annotated using the Trinotate pipeline v3.0.1 (https://trinotate.github.io/). In brief, long Open Reading Frames (ORFs, longer than 100 amino-acids) were first predicted using TransDecoder v3.0.1 (https://transdecoder.github.io/). Long ORFs were queried against Swissprot and Pfam-A [39] databases (last accessed in January 2017) using the Basic Local Alignment Search Tool p + or BLASTp + algorithm (blast+ 2.2.30) [40] and HMMER 3.1b2, [41] respectively. Based on the results from BLAST and Pfam search, TransDecoder predicted proteins from each cluster. If different proteins were predicted for a single cluster, only the best (in terms of homology and sequence length) was retained. These predicted proteins were used as queries for diverse analyses as follows: BLASTp and BLASTx against the Swissprot databases, Pfam-A search, as well as prediction of signal peptide using SignalP v4.1 [42] and transmembrane domains with TMHMM v2.0 [43]. Additionally, clusters and predicted proteins were subjected to BLASTx and BLASTp respectively against the coding sequences and proteins of model

ACCEPTED MANUSCRIPT

species *M. truncatula* v4.0 [44] and *G. max* Wm82.a2.v1, [45] as well as the newly published data from *Lupinus angustifolius* [19]. Trinotate was used to generate a SQLite database containing annotation data for each transcriptomes (thereafter referenced as nodulomes, representing the genes set expressed in nodules and inoculated roots).

2.4. Identification of genes differentially expressed

Prior to expression analysis, low-count clusters were removed from the Corset count matrix using a custom Python script. Low-count clusters were defined as those with less than 20 reads on average in the two conditions, as well as clusters with less than 20 reads in at least two replicates in both conditions. Cleaned count matrices were then normalized using the EDASeq package 2.8.0 [46]. Both between-lanes (library size) and intra-lane (GC content) normalizations were achieved following the full quantile normalization method that was more suitable than other normalization methods (data not shown). Analysis of gene expression was performed using DESeq2 1.14.1 [47] with an adjusted p-value (Bonferroni method) of 0.05. To be considered as differentially expressed (DE), genes should have a p-value less than 0.05. In DESeq2, the incompatible situation between lupine and *Bradyrhizobium* (hereafter named nod^-) was set as reference for the pairwise comparisons with the compatible one (nod^+/nod^-) for each of the three lupine species. Consequently, down-regulated genes are up-regulated with the incompatible bacteria strains, whereas up-regulated genes are up-regulated with the compatible *Bradyrhizobium*.

2.5. Gene expression validation

Gene expression levels were validated by RT-qPCR. Synthesis of cDNA was conducted using iScript™ (BioRad) in 20µL reactions containing 650 ng of RNA. cDNA synthesis was performed following the manufacturer's instructions with an elongation time of 45 min instead of 20 min. Primers were designed for 18 genes belonging to various functional categories using Primer 3 v1.1.3 [48]. Three housekeeping genes, *CBP20*, *PDF2* and *UBQ5*, were also used. Primer sequences and characteristics are described in the Supplementary Table S1. The qRT-PCR was performed using a LightCycler LC480 from Roche (Switzerland) and the accompanying LightCycler® 480 software (version 1.5.1.62) for results visualization and analysis. Each qRT-PCR contained 1µM of cDNA, 0.5µM of each forward and reverse primers and 5µL of Master Mix PowerUp™ SYBR™ Green (ThermoFischer Scientific, USA). The qRT-PCR program was set as follows: an initial activation step at 50°C for 2 min, a first denaturation step at 95°C for 2min following by 45 cycles of denaturation at 95°C for 15s, primer annealing at 56°C for 15s, extension at 72°C for 1min before a dissociation step composed by 15s at 95°C, 1min at 60°C and 15s at 95°C. For relative quantification, the comparative cycle threshold (Ct), also called $2^{-\Delta\Delta Ct}$ was used and the geometric mean of the housekeeping genes employed as reference [49,50].

2.6. Functional analysis of differentially expressed genes

Lupine is a non-model organism and thus results of BLASTx between lupine clusters and the model legume *M. truncatula* proteins performed during the annotation process were used for the functional

ACCEPTED MANUSCRIPT

classification. For each species, the best BLAST hit (according to the highest bitscore, lowest e-value, highest alignment length and identity) was extracted for each of the retained DE genes. To enable the cross-species comparison, lupine DE clusters were grouped according to their best BLAST hit against *M. truncatula*. Reference IDs of *M. truncatula* were then loaded into MapMan software v3.5.1R2 [51] and functional annotations extracted for each species individually. To ensure an accurate analysis, the annotation of genes and metabolic pathways investigated in this work were manually curated. Additionally, an overrepresentation analysis was performed in PageMan as implemented in MapMan [51,52]. The frequency of functional categories in the annotated DE clusters were compared to their expected frequency in background lupine nodulomes (clusters annotated by best BLAST hit from *M. truncatula*) using a Fisher's exact test, an ORA cutoff of 1 and a multiple testing correction of Benjamini-Hochberg.

Normalized counts of DE genes from the corresponding matrix between the lupine species were subjected to a PCA using the ade4 v1.7.8 R package [53] in R v3.4.0 [54] and the graphic interface RStudio v1.0.143 [55]. Visualisation of PCA results was achieved using the R package *ggplot2* v2.2.1 [56].

3. Results and Discussion

3.1. Lupine nodulomes sequencing and assembly

Illumina sequencing of the RNA obtained from nodules (or roots) yielded a total amount (all conditions and replicates merged) of around 215, 224 and 231 million PE reads for *L. albus*, *L. luteus* and *L. mariae-josephae*, respectively (deposited under the NCBI BioProject PRJNA394259). Reads were assembled to produce three transcriptomes that have been obtained for lupine nodules for the first time in the context of symbiotic specificity. The *de novo* assembly of reads with Trinity [36,57] generated between 170,622 and 192,835 contigs. Trinity contigs were then clustered using Corset [37] resulting in 119,403, 132,380 and 130,322 clusters for *L. albus*, *L. luteus* and *L. mariae-josephae*, respectively. For each cluster, the longest contig was considered as representative resulting in a set of sequences of 720-750 bp average length with a N50 value comprised between 1,033 and 1,121 bp and a length spread from 112 to 15,541 bp.

Clusters were subjected to BLASTx and BLASTp against the respective genomes and proteomes of *M. truncatula*, *G. max* and *L. angustifolius* as well as against the SwissProt database. On average, BLAST analyses allowed the annotation of 35% of the clusters (41,618-45,564 clusters), whereas 23-25% were annotated through the Pfam analyses. In addition, all the predicted proteins (32,756-35,232 depending on the species) were successfully annotated using the different BLAST approaches. This is consistent with the roughly 33,000 proteins that have been predicted for *L. angustifolius* [19]. All annotation results were loaded into Trinotate producing easily readable annotation tables (Supplementary tables S2 to S4). Additionally, analysis of the transcriptome completeness was performed using BUSCO [38]. Results revealed that around 70% of the BUSCO groups searched were completely retrieved in the three transcriptomes and almost 90% of them corresponded to single copies of BUSCO groups. The remaining clusters were assigned to fragmented BUSCOs (approximately 11 % for each species) and around 18% of the BUSCOs were absent from all the transcriptomes. These new data come to enrich the growing *Lupinus* database that is mainly constituted of one *Lupinus albus* transcriptome [20], ESTs from *L. luteus* [21] as well as three *L. angustifolius* transcriptomes and its newly published draft genome [19]. In addition, our

ACCEPTED MANUSCRIPT

results provide the first transcriptomic resources for *Lupinus mariae-josephae*, which represent the raw material for further investigations on this enigmatic endemic species [58] that displays a noteworthy symbiotic specificity [25].

3.2. RNA-Seq gene expression analyses and qPCR validation

In order to perform expression analysis, the expression matrix obtained from Corset was filtered to remove low count clusters using a custom Python script as described in section Methods, reducing the size of datasets to around 40,000 clusters for each species. For the differential expression analysis, DESeq2 was used and populated with the EDASeq-normalized data. It identified 1,206, 4,790 and 4,582 differentially expressed (p-value adjusted with the Bonferroni method <0.05) clusters for *L. albus*, *L. luteus* and *L. mariae-josephae*, respectively. Expression levels of 18 genes involved in the RNS and belonging to various functional categories were validated by qPCR (Supplementary Table S5) and more than 80 % of them showed expression patterns consistent with RNA-Seq results. Finally, we applied a fold-change cutoff (1<log2FC <-1) to select the DE genes (1,146, 4,759 and 3,978, respectively) considered for functional analyses (Table 1). Interestingly, *L. albus* exhibited the lowest number of differentially expressed clusters, as compared to *L. luteus* and *L. mariae-josephae*. This is consistent with the ability of this lupine to form nodules with both compatible and incompatible *Bradyrhizobium* strains (even if nodules were inactive with *B. valentinum*), while the two other lupines were only nodulated by their respective compatible species.

Annotated differentially-expressed clusters were grouped according to their *M. truncatula* ID, resulting in a matrix containing 4,103 reference genes. These DE genes are summarized in Supplementary Table S6 including the name and ID of *M. truncatula* orthologs, their expression level in the three lupine species, their MapMan classification and description, as well as the abbreviated and extended names when appropriate. Accordingly, 602, 2,679 and 2,141 DE clusters were assigned to a *M. truncatula* gene ID for *L. albus*, *L. luteus* and *L. mariae-josephae*, respectively. A cross-species comparison of shared and species-specific DE genes (Figure 1) revealed only 86 genes shared by the three lupine species. Among them, 29 were up-regulated and 10 were down-regulated in the three lupines. The closest expression profile was found between *L. luteus* and *L. mariae-josephae*, which shared 774 DE genes, 88% of which had the same expression pattern (Figure 1). On the other hand, *L. albus* shared 95 and 80 DE genes with *L. luteus* and *L. mariae-josephae*, respectively, and around one-half exhibited opposite expression patterns (Figure 1). These results suggest that, at the molecular level, the compatible and incompatible symbiotic interactions within *L. albus* are closer than they are within the two other *Lupinus* species. These contrasted responses of lupine species to compatible or incompatible *Bradyrhizobium* strains were also characterised at the functional and morphological levels by [25]. Indeed, they reported that reduced nodule formation and nitrogen fixation rate occurred between *L. albus* and *B. valentinum* isolate LmjC, indicating the presence of bacteria inside nodules. This interaction represents an intermediate situation between opposite fully compatible or incompatible situations such as those existing in the *L. mariae-josephae* and *L. luteus* systems studied here.

In order to visualize similarities between the different samples and the variables related to that, we performed Principal Component Analysis (PCA) using normalized read counts of each treatment replicate

ACCEPTED MANUSCRIPT

for the 4,103 reference genes defined above (Supplementary Table S6). Compatible and incompatible situations for *L. luteus* and *L. mariae-josephae* were discriminated along the first PCA component (26% of the explained variance; Supplementary Figure S2). Conversely, only a slight differentiation between the two situations was identified for *L. albus* consistently with the observed nodulation in both conditions. In addition, a clear discrimination of the species according to the second and third components (25 and 21% of the explained variance respectively) was also observed, reinforcing the hypothesis of species-specific responses to the bacteria. Finally, analysis of the genes contribution to these three axes revealed that none of the 4,103 genes contributed more than 1% of the axes variance, suggesting that responses to compatible or incompatible *Bradyrhizobium* are driven by an important set of genes, thus illustrating the complexity of the interaction.

3.3. Overview of the functional classification of nodulomes and differentially expressed clusters

To assess the functional categories represented in the three lupine transcriptomes, BLASTx results against the *M. truncatula* proteome (Supplementary Tables S2-S4) were used in order to assign, when applicable, a *M. truncatula* ID correspondence to each cluster. Functional classification of the lupine transcriptomes as well as of the DE clusters assigned to a *M. truncatula* ID were performed using the MapMan software [51]. Since the MapMan software does not handle duplicated IDs, in case of no one-to-one correspondence the best BLAST hit (defined as in section Methods) was kept for each reference gene. This produced a final matrix with around 40 % of the clusters (*ca.* 16,000 genes) that was used as input for the MapMan software. Cross-species comparison results revealed that among the assigned clusters (15,928 to 16,504 according to the species), more than 75% (12,512) were common to the three lupine species, whereas only around 8% of each transcriptome was species-specific (between 1,240 and 1,532; Figure 2). This result suggests that most of the genes expressed in nodules and roots in compatible or incompatible situations are shared by the three species. Concerning functional categories at the transcriptome scale, all the MapMan bin categories are similarly represented in the three *Lupinus* species (Supplementary Figure S3), in accordance with the above result.

Regarding the functional classification of the DE genes for each lupine species, up- and down-regulated ones were discriminated and separately treated in order to draw a more detailed understanding of the expression patterns across different functional categories. To this end, functional category enrichment tests were conducted for each species using PageMan [51,52]. Different functional categories were over-represented in both up- and down-regulated genes of either *L. albus*, *L. luteus* and *L. mariae-josephae* (Figure 3). However, the two latter species share more enriched functional categories. Thus, functional enrichment results for *L. albus* revealed only three functional categories (‘Misc’, ‘Hormone metabolism’ and ‘Minor CHO metabolism’) containing over-represented functions in up-regulated genes only (Figure 3A). This is again consistent with the fact that *L. albus* produced nodules with both bacterial species. On the contrary, 13 and 10 enriched functional categories were found in *L. luteus* and *L. mariae-josephae*, respectively. For these two species, ‘Hormone metabolism’, ‘Misc’, and ‘Transport’ contained over-represented functions in both up- and down-regulated genes, whereas functions from the ‘Signalling’, ‘RNA’, ’Stress’ and ‘Secondary metabolism’ categories were over-represented only among down-regulated

ACCEPTED MANUSCRIPT

genes (Figure 3B and 5C). All other over-represented functional categories ('Lipid metabolism', 'major CHO metabolism', 'Amino-acid metabolism', 'Glycolysis', 'Redox', 'N-metabolism', and 'Protein') either occurred in only one of these two lupine species or presented an opposite regulation of expression in the two species (Figure 3B and 5C). All in all, these analyses notably pointed out the pathways related with hormone metabolism, CHO metabolism, amino-acid and N-metabolism, stress and secondary metabolism as being part of the symbiotic response to different *Bradyrhizobium* strains in lupine roots. Genes belonging to similar functional categories were also found significantly affected during the RNS in the model legume *M. truncatula* [59] and references therein; [60]. In addition to these categories, we also retained for further analysis the canonical symbiotic genetic pathway, since it is not represented by a defined MapMan functional category but involves genes from diverse functional categories. The way all these different, functionally-related genes can help explain the different mechanisms implicated in the host-symbiont choice and how they participate in the establishment of a successful symbiosis has been carefully analysed in the following sections.

3.4. Most of the canonical symbiotic genes are DE in both *L. luteus* and *L. mariae-josephae*

Root nodule symbiosis is regulated by a complex genetic pathway involving genes from diverse functional categories (reviewed in [7,8]). All of these canonical symbiotic genes, previously identified in model legumes and in the genome of *L. angustifolius*, were retrieved in the three lupine transcriptomes analysed in this study. Interestingly, they were not differentially expressed in *L. albus* irrespective of the bacterial symbiont (Figure 4). Although this is consistent with the formation of nodules with both *Bradyrhizobium* strains, it also suggests that the infection process was successful in both cases, since genes controlling this step, such as *CCamk*, *CycloPS*, *NSP1*, *NSP2* and *ERN1* did not exhibit significant expression changes in this species. A similar situation has been observed in *M. truncatula*, where the ecotype Jemalong 6 showed non-functional infected nodules with *Sinorhizobium meliloti* strain A145 [32]. Thus, the non-functionality of *L. albus* nodules could be due to an incompatibility occurring in the later stages of symbiosis such as the nitrogen and carbohydrates exchanges, or to a late activation of plant defence. Conversely, genes expressed at different steps of the RNS and in both epidermal and cortical programs were significantly affected by the change of symbiont in both *L. luteus* and *L. mariae-josephae* (Figure 4). Although lupine orthologs of the NOD factor receptors *NFR1* and *NFR5* were not DE in these species, the expression of other transmembrane receptors (i.e. *SymRK*, *EPR3* and concavanalin A-like lectin genes) involved in RNS was significantly higher in functional nodulation situations. As an explanation for the low expression of other receptors (e.g. *NFR1* and *NFR5*) for rhizobia effectors, it must be kept in mind that nodules used here were mature and not at the early stages of plant-rhizobia molecular dialogue, when these receptors are involved [10]. Alternatively, even if still involved at this stage of the RNS, it is possible that regulation occurs on the other components of the genetic pathway.

Further downstream of the symbiotic signalling cascade (Figure 4), several key genes belonging to epidermal and/or cortical programs were differentially expressed and exhibited a significant up-regulation in both *L. luteus* and *L. mariae-josephae*. This is the case of key genes *CCamK* and *CycloPS*, which encode proteins acting downstream of the calcium spiking with a central role in the coordination of epidermal and

ACCEPTED MANUSCRIPT

cortical programs, and the *MCA8* gene, encoding a cation channel required for calcium spiking. During the RNS, CCAMK and CYCLOPS form a complex that is thought to be able to activate transcription factors such as *NSP1*, *NSP2* or *NIN*. Neither *NSP1* or *NSP2* showed significant change of expression in this experiment, whereas *NIN* was induced by the compatible bacteria in both *L. luteus* and *L. mariae-josephae* [63]. Interestingly, other recent findings revealed that CYCLOPS is also able to directly activate the transcription of *NIN*, which is also essential for the RNS. Thus, these different transcription factors downstream of CYCLOPS appear to be involved in RNS and are known to activate nodulation genes such as *ENOD11*, *RPG*, or *ERN1* (the latter being up-regulated in both *L. luteus* and *L. mariae-josephae*) [63]. Finally, two more genes (*PUB1* and *VAPYRIN*) implicated in epidermal programs were also induced in *L. luteus* and *L. mariae-josephae*. *PUB1* (Plant U-box protein 1) is known to encode a protein able to interact with and to negatively regulate the NOD factor receptor *LYK3* (ortholog of *NFR1*) in *M. truncatula* and could play a role in the symbiotic specificity [62]. The *VAPYRIN* gene is also essential for the establishment of both mycorrhizal symbiosis and RNS in *M. truncatula* [63].

Once roots infected, bacteria are released as symbiosomes into the cytoplasm of infected cells, where they undergo elongation and differentiation stages, resulting in functional bacteroids. Several genes involved in these steps were identified in the DE dataset, such as the remorin-like encoding gene *SYMREM1* and *RSD* involved in the bacteria elongation regulation [64] (up-regulated in both *L. luteus* and *L. mariae-josephae*), or *NAC074* (NAC transcription factor) which negatively affects symbiosome [65] and was down-regulated in *L. luteus* and *L. mariae-josephae*. Finally, two transcription factors regulating the bacteroid development and differentiation (*NF-YA1* and *EFD*) were up-regulated in *L. luteus* and *L. mariae-josephae*, which is in line with bacteroid differentiation. It is worth noting that despite the implication of numerous genes in lupine nodule formation, several RNS canonical genes (e.g. the nucleoporins *NUP133*, *NUP85*, *NENA* as well as the cation channels *CASTOR* and *POLLUX*) were not significantly affected by the symbiont switch. This expression pattern could be explained by temporal variations of gene expression during nodulation, since nodules from different stages were used. Such temporal gene expression variations have already been demonstrated on soybean and barrel medic [59,60]. Canonical symbiotic genes are significantly induced by compatible symbionts in both *L. luteus* and *L. mariae-josephae*, suggesting an activation of these genes in compatible situation. However, they are generally expressed in early steps of the symbiotic interaction in model legumes, and thus, this needs to be further confirmed during the first steps of the *Bradyrhizobium*/lupines interaction.

3.5. Carbon and nitrogen metabolism is induced in compatible situations

Root nodule symbiosis induces significant changes in carbon and nitrogen partitioning in the host plant. Fixed carbon from photosynthesis (converted into malate) is provided to the nitrogen-fixing bacteroids and fixed nitrogen is assimilated by the plant host [1]. Genes related to these processes were overall up-regulated in the three lupine species. Starch biosynthesis was globally up-regulated in both *L. luteus* and *L. mariae-josephae*. Thus, genes coding for three starch synthases, one granule-bound starch synthase and one starch-branching enzyme were up-regulated in these species (Figure 5, top left). Interestingly, genes for

ACCEPTED MANUSCRIPT

starch catabolism were also up-regulated, with different α- and β-amylases up-regulated in the three lupine species, as well as a starch- degradation enzyme-encoding gene (StP) in *L. mariae-josephae* (Figure 5, top left). This is in accordance with similar results obtained in *L. japonicus* [66]. Indeed, accumulation of starch in nodules is reduced and has only been observed in defective nodules which are not a strong C sink [66].

Carbohydrates derived from photosynthesis are the source of bacteroid fuel for nitrogen fixation and they are transported from leaves to nodules in the form of sucrose [67]. In nodules, sucrose is converted to fructose and glucose or UDP-Glucose by invertases and sucrose synthases respectively. Interestingly, these enzymes, encoded by multigene families, displayed a temporal expression variation in nodules of model legumes: invertases are generally induced in the developing nodules, whereas sucrose synthases are more active in mature nodules [67,68]. Since nodules from different development stages were used in this study, this could explain why five invertases and five synthases were significantly affected by the symbiont switches. Most of them were down-regulated in either *L. luteus*, *L. mariae-josephae* or in both species, except for two sucrose synthases and one invertase diversely up-regulated. Interestingly, these two up-regulated sucrose synthases are known to be strongly induced in *M. truncatula* nodules according to the *Medicago truncatula* genome database [69].

Fructose, glucose and UDP-glucose produced by invertases and sucrose synthases are precursors involved in the biosynthesis of different sugars such as raffinose, galactose, trehalose or myo-inositol. Interestingly, genes involved in the biosynthesis of these sugars were mainly up-regulated in the three species (Figure 5). Although trehalose is mainly produced by the bacteria and is thought to act as an osmoprotectant, essential during infection [70], previous studies highlighted that legume-originated trehalose has a significant, positive impact on the nodule size, structure and nitrogen fixation, as well as on the resident bacteroids structure, number and metabolism [71,72]. Accordingly, the up-regulation of trehalose biosynthesis in the three lupine species supports the hypothesis of a functional symbiosis in compatible situations. Fructose is also the starting point of the glycolysis resulting in the production of malate provided to the bacteroid (Figure 5). The transformation of fructose 6P into fructose 1,6BP is achieved by a phosphofructokinase, which is up-regulated in *L. luteus* and *L. mariae-josephae.* Subsequent steps lead to the production of phosphoenol pyruvate (PEP), then converted to oxaloacetate (OAA) by PEP carboxylase (PEPC), which is up-regulated in the three lupine species. The OAA is then turned into malate by the malate dehydrogenase enzyme, up-regulated in *L. luteus* only.

Once fixed by bacteroids, nitrogen is exported (as NH_4^+) through the symbiosome membrane and assimilated by the legume host. Ammonium cation (NH_4^+) is assimilated in the glutamine/glutamate cycle (Figure 5, lower right ; [73]). Five GS-encoding genes were differentially expressed and mainly up-regulated in *L. luteus* or *L. mariae-josephae*. Conversely, one *GOGAT*, as well as three asparagine synthase (AS) encoding genes were significantly up-regulated in the three lupine species, whereas the asparagine aminotransferase (AAT) gene was up- and down-regulated in *L. albus* and *L. mariae-josephae*, respectively. The strong up-regulation of key genes involved in the glutamine/glutamate cycle supports the hypothesis of a functional assimilation of fixed nitrogen with compatible *Bradyrhizobium* strains and

ACCEPTED MANUSCRIPT

reinforce the hypothesis that there is no functional exchange of carbon/nitrogen in the *L. albus / B. valentinum* relationship. Finally, it is interesting to note that different genes related to nitrogen foraging, transport and partitioning (such as diverse *NRT* and *AMT*) were found to be down-regulated in both *L. luteus* and *L. mariae-josephae* suggesting that a nitrogen-stress response is triggered in the incompatible situations [74]. The functionality of *L. luteus* and *L. mariae-josephae* with their respective compatible symbiont is also supported by up-regulation of leghaemoglobin encoding genes for which the *M. truncatula* homologs are highly expressed in nodules under symbiotic conditions [69]. Additionally, two homologs of *VIT1* and *NRAMP3* genes, which are essential for iron transport in the symbiosome [75], were retrieved among the up-regulated *L. mariae-josephae* genes. It is known that iron is involved in symbiotic N_2 fixation, and positively affected nodule initiation and function in lupines [76].

3.6. Hormone biosynthesis and signalling pathways are strongly affected in a symbiont-dependent manner

Hormones play an essential role in diverse aspects of the plant life cycle, and molecules such as ethylene, auxins or cytokinins, are key components of the nodule organogenesis and infection process regulation [77]. 'Hormone metabolism' MapMan functional category was over-represented in the three lupine species studied here (Figure 3), comprising 180 DE genes which were mainly assigned to the ethylene and auxin biosynthesis pathways (each hormone represented 20 % of the classified genes). This is consistent with recent findings that highlight the central and complex role of ethylene at different stages of the RNS [65,78], which also is underlined by the over-representation of ethylene signalling sub-categories observed among the down-regulated DE genes in both *L. luteus* and *L. mariae-josephae* (Figure 3). In addition, gibberellins, cytokinins, abscisic acid and jasmonate metabolisms represent between 5 and 15 % of the genes classified in the 'Hormone metabolism' category.

Recently, a detailed survey of the ethylene impact on RNS revealed a dual role for this hormone. On one side, ethylene acts as a negative regulator of RNS in a NOD-dependent fashion and as an inducer of plant defense-related genes. On the other side, ethylene is able to induce the expression of several genes involved in the nodulation process [65]. Indeed, genes encoding the two key enzymes involved in ethylene biosynthesis, i.e. 1-aminocyclopropane-1-carboxylic acid synthase (ACS) and 1-aminocyclopropane-1-carboxylic acid oxidase (ACO) were up-regulated in *M. truncatula* during symbiosis with *S. meliloti* [78]. These genes were significantly affected by the compatibility between lupine hosts and *Bradyrhizobium* strains. Four *ACS* genes displayed contrasting expression patterns in both *L. luteus* and *L. mariae-josephae*, whereas the only *ACS* gene DE in *L. albus* was down-regulated. Conversely, one ACO encoding gene was significantly induced by the incompatible *Bradyrhizobium* strains in both *L. luteus* and *L. mariae-josephae*. These results suggest different regulation states of ethylene biosynthesis that could be explained by the dual role played by this hormone, or by the action of two regulation systems for ACS or ACO. Ethylene signalling was also highly affected and generally down-regulated in both *L. luteus* and *L. mariae-josephae* suggesting an activation of plant defences in reaction to the incompatible bacterial (discussed later). Several genes coding for AP2/Ethylene-responsive elements binding proteins were down-regulated in both *L. luteus* and *L. mariae-josephae*, with the exception of the essential symbiotic *ERN1* gene [78,79] which is

ACCEPTED MANUSCRIPT

up-regulated in *L. luteus*. Taken together, these results strongly suggest that the incompatible *Bradyrhizobium* strain induced ethylene-driven defence responses in both *L. luteus* and *L. mariae-josephae*, whereas the positive role of ethylene during the RNS was less detectable.

The balance between auxins and cytokinins is an essential regulator of nodule organogenesis. Previous studies revealed that nodule initiation requires low auxin and high cytokinins levels. Once nodules are initiated, auxin levels are increased and auxin-responsive genes are induced in the nodule meristem [77,80]. Investigation of auxin-related genes revealed that lupine homologs to *MtTAR2* and *YUC* genes, involved in the biosynthesis of the indole 3-acetic acid (IAA), are mainly up-regulated in either *L. luteus*, *L. mariae-josephae* or both, which was also shown for *YUC* genes in *M. truncatula* [78]. During RNS, auxin levels in nodule primordia are regulated through the synthesis/degradation of the auxin transporters or through the action of different transport regulators [77]. Thus, in general, auxin transporters (*PIN*, *ABCB15*, *ABCB19*, *LAX2*) as well as their regulators (such as *PINOID*, *PIPK5*, *CHMP1A*, *NPY1*, *IAR3*, *IAMT1* or *AFB*) were mainly up-regulated in both *L. luteus* and *L. mariae-josephae* although some of them displayed contrasting expression patterns. Finally, among auxin responsive elements known to be rapidly induced by auxins and down-regulated during nodule initiation [80], different expression patterns were identified in both *L. luteus* and *L. mariae-josephae*, suggesting that auxins are subjected to spatial and temporal variations as it has been previously revealed in model legumes.

Cytokinins are known to be essential for nodule morphogenesis initiation and mediate the localized down-regulation of auxin transport genes during this step [77]. Here, the cytokinins biosynthesis pathway was found to be mainly up-regulated in the species nodulated by only one *Bradyrhizobium* strain (Supplementary Figure S4). Indeed, three ITP encoding genes, required for the isopentenyladenine and the *trans*-Zeatin biosynthesis, were found up-regulated in either *L. luteus* or *L. mariae-josephae*. These genes are classically up-regulated during nodule morphogenesis in *M. truncatula* [81]. The cytokinin signalling pathway was found globally induced by the compatible symbiont in *L. luteus* and *L. mariae-josephae*. Although cytokinin receptors (called AHK) had divergent expression patterns in lupines, their target genes, involved in the regulation of cytokinins biosynthesis (*LOG* genes), were mainly up-regulated in both *L. luteus* and *L. mariae-josephae* as it was shown during *M. truncatula* symbiosis [82]. Similarly, AHP proteins, which are phosphorylated by AHKs receptors, were also up-regulated in both *L. luteus* and *L. mariae-josephae*. These results are consistent with nodule development in compatible situations. AHPs signal is transmitted to A and B-type response regulators (ARRs) which determinate the cytokinin output [83]. Only A-type ARRs were retrieved and they were up-regulated in both *L. luteus* and *L. mariae-josephae*, whereas one was down-regulated in *L. albus*. One of them (the *RR4* gene) is a negative regulator of nodule number induced during the *M. truncatula* symbiosis, partly activated by an ethylene response factor required for nodule differentiation (*EFD*) [84]. Thus, these results suggest the activation of cytokinin biosynthetic genes, although the signalling pathway presents a more contrasting expression pattern that could be due to temporal and/or spatial variations in cytokinin signalling (as shown in [85]).

Similarly to ethylene, abscisic acid (ABA) is a negative regulator of nodulation and is able to regulate NOD factors signalling, cytokinin-dependent nodule initiation as well as calcium spiking signature [86]. ABA-related genes revealed an up-regulation of two *NCED* and one *ABA4* genes, required for the ABA

ACCEPTED MANUSCRIPT

biosynthesis, whereas the different ABA-responsive genes were down-regulated in the three lupine species. Indeed, one *PYR* and three *PYL* genes encoding ABA receptors were down-regulated in *L. luteus* and *L. mariae-josephae*. Under ABA activation, PYL receptors inhibit phosphatase 2C protein (PP2C, four genes identified and down-regulated in the three species) resulting in the induction of ABA responsive genes [87]. Although most of these genes have been described in response to abiotic stress, such as drought [88], our results suggest the activation of the ABA responsive genes in the incompatible situations. This is consistent with the fact that ABA is thought to act as an antagonist of cytokinins, resulting in an inhibition of nodulation [77].

Genes coding for enzymes of the gibberellins (GAs) biosynthetic pathway (*GA1, GA2, KAO, GA20ox*, and *GA3ox*) were in general up-regulated in *L. luteus* and *L. mariae-josephae* (Figure 6). GAs can be inactivated through the GA2ox activity (one gene up-regulated in *L. luteus* and another down-regulated in *L. luteus* and *L. mariae-josephae*). The up-regulation of the GAs biosynthesis pathway observed here in both *L. luteus* and *L. mariae-josephae* has been also described in other legumes during their nodulation [89–91]. However, GAs are known to inhibit nodulation [92] and the downstream signalisation pathway revealed the activation of GAs signal transduction repressors. Indeed, biologically active GAs induce the transduction of a signal through the GID1 receptor (the coding gene is down-regulated in both *L. luteus* and *L. mariae-josephae*) promoting the degradation of DELLA proteins which suppress the transcription of the GA-responsive genes [92]. DELLA proteins are essential for the RNS and have been recently found to promote the formation of the CCAMK/CYCLOPS complex and bridging it with NSP1/NSP2 [92]. Here, one *DELLA* gene was up-regulated in *L. luteus* and *L. mariae-josephae*. In addition, DELLA-mediated repression of GAs is also supported by the up-regulation, in *L. luteus* only, of *BRG3*, a S-ribonuclease binding protein encoding gene, required for a complete GA repression by DELLAs [93]. Taken together, these results suggest that the presence of a compatible *Bradyrhizobium* strain induced GAs biosynthesis in both *L. luteus* and *L. mariae-josephae*, whereas the GA signalling pathway through the GID1 receptor was repressed. Consequently, this suggests a fine control of the GAs concentration as it has been shown in *Pisum sativum*, where an optimal concentration of GAs was required for proper RNS [89].

3.7. Flavonoid biosynthesis pathway is induced by incompatible strains in *L. luteus* and *L. mariae-josephae*

'Secondary metabolism' category was one of the four functional categories over-represented only in down-regulated genes of *L. luteus* and *L. mariae-josephae* (Figure 3B-C). Among this class we notably found flavonoids which are major compounds present in all vascular plants and involved in a huge range of plant life aspects [94]. Legumes are particularly rich in (iso)flavonoids and there is no doubt about their central role during RNS [95,96]. Indeed, they act as either specific inducers or repressors of rhizobia growth, inducers of NOD-factors biosynthesis and regulators of the auxin transport required for nodule organogenesis [96,97]. From phenylalanine to the different flavonoids, most of the identified genes coding for biosynthetic enzymes (i.e. PAL, C4H, 4CL, CHS, CHI, F3'H, FLS, FNS) were significantly down-regulated in both *L. luteus* and *L. mariae-josephae* (Figure 7). Only three biosynthetic genes appeared to be up-regulated, namely *C4H*, *CHS* and *F3H* genes. Interestingly, no genes encoding key enzymes known to

ACCEPTED MANUSCRIPT

catalyse the biosynthesis of isoflavonoids were identified among the DE genes. This is surprising since previous studies showed the accumulation of diverse isoflavonoids in nodules of *L. albus* and *L. luteus* [98,99]. However, it is well known that (iso)flavonoids can play species-specific roles and the activation of flavonoid biosynthesis in compatible situations could be masked by a higher induction of these pathways under incompatible interactions for nodulation [94].

3.8. Plant defences are induced by the incompatible *Bradyrhizobium* strains

The 'Stress' functional category was another of the over-represented functional categories among down-regulated genes in *L. luteus* and *L. mariae-josephae* (Figure 3B-C). This is of particular interest in the context of biotic interactions and is consistent with a defence response against the incompatible *Bradyrhizobium* strains.

The first active response of plants to microbes is the primary immune response (PTI=Pathogen Triggered Immunity), which is initiated upon perception of conserved microbial molecules by plant cell-surface receptors [16]. Some of these microbial features are proteinaceous elicitors, such as the bacterial flg22 from flagellin, that are perceived by receptors with LRR-type extracellular domains. PTI leads to the induction in plants of pathogen-responsive genes and defences through a signalling pathway including ROS, hormones and MAP kinases, among other. Different genes related to these mechanisms were mainly induced in the *L. luteus*/LmjC and *L. mariae-josephae*/ISLU101 non-nodulating interactions. Indeed, different pathogenesis-related proteins (PR1, PR5), the RPM1 protein and some of its downstream targets (such as RIN4, GSTF9, RIPK4), four *ERF1* transcription factors, as well as the flg22 receptor *FLS2* and its associated genes (e.g *PBL2*, *MAPKK5*, *PUB17* and *SOBIR1*), were retrieved among the down-regulated genes of *L. luteus* and *L. mariae-josephae*. In addition, genes involved in the MAMP-triggered immunity such as *RLK7* or *PUB23* and *PUB24* and in hormonal regulation of defence responses (for example CBP60g, CPK6 and GRX480 promoting the accumulation of salicylic acid (SA) and regulating jasmonic acid (JA) and JA/SA signalling) were also down-regulated in these species. Interestingly, several genes belonging to the lignin biosynthesis and secondary cell wall formation pathways (*i.e.* CAD, HCT and CCoAOMT) were specifically down-regulated in *L. luteus* only. In addition, numerous MYB transcription factors (MYB4, 103, 43, 20, 52, 63) and ß-glucosidase encoding-genes, as well as an ankyrin-repeat family gene involved in the regulation of the lignin biosynthesis or cell wall formation also displayed a similar expression pattern. These genes have been shown to be activated during plant responses to pathogens [100] (and references therein), which suggests that *B. valentinum* likely induced this process in *L. luteus*.

4. Conclusion

In this study, nodule and root transcriptomes of three Old-World lupine species (*L. albus*, *L. luteus* and *L. mariae-josephae*) cultivated with compatible and incompatible *Bradyrhizobium* strains were reconstructed and analysed in order to better understand the host control and mechanisms of symbiotic specificity. This is also the first report of genomic resources for the endemic *Lupinus mariae-josephae*, which represent new data for further studies on this enigmatic species. Analysis of gene expression revealed that *L. luteus* and *L. mariae-josephae* (each restricted to a compatible bacterial strain) had more than twice differentially-

ACCEPTED MANUSCRIPT

expressed genes as compared to *L. albus* (nodulating with both bacterial strains, but functional with only one). In addition, these genes presented more similar expression pattern between *L. luteus* and *L. mariae-josephae* than between these two species and *L. albus*. This indicates that the two former species exhibit similar responses to the compatible or incompatible bacteria. Functional classification of the DE genes showed that most of the canonical symbiotic genes (involved in both epidermal and cortical programs) were significantly induced by the compatible bacteria, whereas none of them was differentially expressed in *L. albus*, suggesting that the infection of *L. albus* nodules by *B. valentinum* occurred successfully. This also indicates that the incompatibility between *L. albus* and *B. valentinum* was determined at later stages of the RNS. This was also supported by a detailed analysis of the carbon and nitrogen metabolism, in which genes related to the fixed nitrogen (e.g., asparagine synthase, GOGAT) and to the carbohydrates supply to bacteroids (PEPC, for example) were strongly induced in the three lupines by their compatible *Bradyrhizobium* strains. Other functional categories were also significantly affected by the symbiont switch, such as defence response, hormone and secondary metabolisms. Our results suggest an activation of similar defence mechanisms (ethylene-related defence genes, *RPM1* and *FLS2* signalling pathway) in both *L. luteus* and *L. mariae-josephae* when challenged with the incompatible species, but also more specific processes (secondary cell wall formation in *L. luteus*). This first study conducted in the genistoid legume genus *Lupinus* provides a first insight on the regulation mechanisms underlying interactions between lupines and their nitrogen-fixing symbionts, and allowed circumscription of a consistent set of candidate genes potentially involved in symbiotic host-specificity. Our results open new avenues to more accurately investigate the role of these genes (or gene networks) throughout the different nodulation steps. Indeed, further consideration and comparative analyses of gene expression at different stages of the plant/bacteria association will most likely improve our understanding on symbiotic specificity.

Funding

JK was supported by a doctoral research grant from the University of Rennes 1 – French Ministry of Higher Education and Research. This work benefited from the International Associated Laboratory "Ecological Genomics of Polyploidy" supported by CNRS (INEE, UMR CNRS 6553 Ecobio), University of Rennes 1, Iowa State University (Ames, USA). Work at CBGP was supported by MICINN (CGL-26932 to JI).

Conflict of interest

None declared.

Acknowledgements

We are grateful to the Genouest Bioinformatics (INRIA, Rennes) and Bioinfo Genotoul (Toulouse Midi-Pyrenees) platforms for providing computing resources. The analyses benefited from the Environmental and Human Genomics platform (UMR CNRS 6553 Ecobio) facilities. The authors also thank Jean-Bernard Magné-Robert (INRA UMR 1347 Agroécologie, Dijon, France) for kindly providing seeds of *L. luteus*. We are grateful to Delphine Giraud for her help in the gene expression analysis.

ACCEPTED MANUSCRIPT

References

[1] U.A. Hartwig, The regulation of symbiotic N2 fixation: a conceptual model of N feedback from the ecosystem to the gene expression level, Perspect. Plant Ecol. Evol. Syst. 1 (1998) 92–120.

[2] G.P. Lewis, Legumes of the World, Royal Botanic Gardens, Kew, 2005.

[3] J.I. Sprent, J. Ardley, E.K. James, Biogeography of nodulated legumes and their nitrogen-fixing symbionts, New Phytol. (2017). doi:10.1111/nph.14474.

[4] P.H. Graham, C.P. Vance, Legumes: Importance and Constraints to Greater Use, PLANT Physiol. 131 (2003) 872–877. doi:10.1104/pp.017004.

[5] F. Cabello-Hurtado, J. Keller, J. Ley, R. Sanchez-Lucas, J.V. Jorrín-Novo, A. Aïnouche, Proteomics for exploiting diversity of lupin seed storage proteins and their use as nutraceuticals for health and welfare, J. Proteomics. 143 (2016) 57–68. doi:10.1016/j.jprot.2016.03.026.

[6] G.E.D. Oldroyd, J.A. Downie, Calcium, kinases and nodulation signalling in legumes, Nat. Rev. Mol. Cell Biol. 5 (2004) 566–576. doi:10.1038/nrm1424.

[7] G.J. Desbrosses, J. Stougaard, Root Nodulation: A Paradigm for How Plant-Microbe Symbiosis Influences Host Developmental Pathways, Cell Host Microbe. 10 (2011) 348–358. doi:10.1016/j.chom.2011.09.005.

[8] G.E.D. Oldroyd, Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants, Nat. Rev. Microbiol. 11 (2013) 252–263. doi:10.1038/nrmicro2990.

[9] S.R. Long, Genes and signals in the Rhizobium-legume symbiosis, Plant Physiol. 125 (2001) 69–72.

[10] S. Radutoiu, L.H. Madsen, E.B. Madsen, H.H. Felle, Y. Umehara, M. Grønlund, S. Sato, Y. Nakamura, S. Tabata, N. Sandal, others, Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases, Nature. 425 (2003) 585–592.

[11] S. Radutoiu, L.H. Madsen, E.B. Madsen, A. Jurkiewicz, E. Fukai, E.M. Quistgaard, A.S. Albrektsen, E.K. James, S. Thirup, J. Stougaard, LysM domains mediate lipochitin–oligosaccharide recognition and Nfr genes extend the symbiotic host range, EMBO J. 26 (2007) 3923–3935.

[12] D.W. Ehrhardt, E.M. Atkinson, S.R. Long, Depolarization of alfalfa root hair membrane potential by Rhizobium meliloti Nod factors, Science. 256 (1992) 998–1000.

[13] N. Kanamori, L.H. Madsen, S. Radutoiu, M. Frantescu, E.M. Quistgaard, H. Miwa, J.A. Downie, E.K. James, H.H. Felle, L.L. Haaning, others, A nucleoporin is required for induction of Ca2+ spiking in legume nodule development and essential for rhizobial and fungal symbiosis, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 359–364.

[14] K. Saito, M. Yoshikawa, K. Yano, H. Miwa, H. Uchida, E. Asamizu, S. Sato, S. Tabata, H. Imaizumi-Anraku, Y. Umehara, H. Kouchi, Y. Murooka, K. Szczyglowski, J.A. Downie, M. Parniske, M. Hayashi, M. Kawaguchi, NUCLEOPORIN85 Is Required for Calcium Spiking, Fungal and Bacterial Symbioses, and Seed Production in Lotus japonicus, Plant Cell Online. 19 (2007) 610–624. doi:10.1105/tpc.106.046938.

[15] J. Lévy, C. Bres, R. Geurts, B. Chalhoub, O. Kulikova, G. Duc, E.-P. Journet, J.-M. Ané, E. Lauber, T. Bisseling, J. Dénarié, C. Rosenberg, F. Debellé, A Putative Ca2+ and Calmodulin-Dependent Protein Kinase Required for Bacterial and Fungal Symbioses, Science. 303 (2004) 1361–1364. doi:10.1126/science.1093038.

[16] B. Gourion, F. Berrabah, P. Ratet, G. Stacey, Rhizobium–legume symbioses: the crucial role of plant immunity, Trends Plant Sci. 20 (2015) 186–194. doi:10.1016/j.tplants.2014.11.008.

[17] Q. Cronk, I. Ojeda, R.T. Pennington, Legume comparative genomics: progress in phylogenetics and phylogenomics, Curr. Opin. Plant Biol. 9 (2006) 99–103. doi:10.1016/j.pbi.2006.01.011.

[18] J.S. Gladstones, C.A. Atkins, J. Hamblin, eds., Lupins as crop plants: biology, production, and utilization, CAB International, Wallingford, Oxon, UK ; New York, NY, USA, 1998.

[19] J.K. Hane, Y. Ming, L.G. Kamphuis, M.N. Nelson, G. Garg, C.A. Atkins, P.E. Bayer, A. Bravo, S. Bringans, S. Cannon, D. Edwards, R. Foley, L. Gao, M.J. Harrison, W. Huang, B. Hurgobin, S. Li, C.-W. Liu, A. McGrath, G. Morahan, J. Murray, J. Weller, J. Jian, K.B. Singh, A comprehensive draft genome sequence for lupin (*Lupinus angustifolius*), an emerging health food: insights into plant-microbe interactions and legume evolution, Plant Biotechnol. J. (2016). doi:10.1111/pbi.12615.

[20] J.A. O'Rourke, S.S. Yang, S.S. Miller, B. Bucciarelli, J. Liu, A. Rydeen, Z. Bozsoki, C. Uhde-Stone, Z.J. Tu, D. Allan, J.W. Gronwald, C.P. Vance, An RNA-Seq Transcriptome Analysis of Orthophosphate-Deficient White Lupin Reveals Novel Insights into Phosphorus Acclimation in Plants, PLANT Physiol. 161 (2013) 705–724. doi:10.1104/pp.112.209254.

ACCEPTED MANUSCRIPT

[21] L.B. Parra-González, G.A. Aravena-Abarzúa, C.S. Navarro-Navarro, J. Udall, J. Maughan, L.M. Peterson, H.E. Salvo-Garrido, I.J. Maureira-Butler, Yellow lupin (*Lupinus luteus* L.) transcriptome sequencing: molecular marker development and comparative studies, BMC Genomics. 13 (2012) 425.

[22] A. Gonzalez-Sama, M.M. Lucas, M.R. de Felipe, J.J. Pueyo, An unusual infection mechanism and nodule morphogenesis in white lupin (*Lupinus albus*), New Phytol. 163 (2004) 371–380. doi:10.1111/j.1469-8137.2004.01121.x.

[23] T. Stepkowski, C.E. Hughes, I.J. Law, L. Markiewicz, D. Gurda, A. Chlebicka, L. Moulin, Diversification of Lupine Bradyrhizobium Strains: Evidence from Nodulation Gene Trees, Appl. Environ. Microbiol. 73 (2007) 3254–3264. doi:10.1128/AEM.02125-06.

[24] H. Pascual, " *Lupinus mariae-josephi*" (Fabaceae), nueva y sorprendente especie descubierta en España, in: An. Jardín Botánico Madr., Real Jardín Botánico, 2004: pp. 69–72. http://dialnet.unirioja.es/servlet/articulo?codigo=1033446 (accessed January 26, 2016).

[25] C. Sánchez-Cañizares, L. Rey, D. Durán, F. Temprano, P. Sánchez-Jiménez, A. Navarro, M. Polajnar, J. Imperial, T. Ruiz-Argüeso, Endosymbiotic bacteria nodulating a new endemic lupine *Lupinus mariae-josephi* from alkaline soils in Eastern Spain represent a new lineage within the Bradyrhizobium genus, Syst. Appl. Microbiol. 34 (2011) 207–215. doi:10.1016/j.syapm.2010.11.020.

[26] D. Wang, S. Yang, F. Tang, H. Zhu, Symbiosis specificity in the legume - rhizobial mutualism: Host specificity in root nodule symbiosis, Cell. Microbiol. 14 (2012) 334–342. doi:10.1111/j.1462-5822.2011.01736.x.

[27] Y. Kawaharada, S. Kelly, M.W. Nielsen, C.T. Hjuler, K. Gysel, A. Muszyński, R.W. Carlson, M.B. Thygesen, N. Sandal, M.H. Asmussen, M. Vinther, S.U. Andersen, L. Krusell, S. Thirup, K.J. Jensen, C.W. Ronson, M. Blaise, S. Radutoiu, J. Stougaard, Receptor-mediated exopolysaccharide perception controls bacterial infection, Nature. 523 (2015) 308–312. doi:10.1038/nature14611.

[28] J. Liu, S. Yang, Q. Zheng, H. Zhu, Identification of a dominant gene in Medicago truncatula that restricts nodulation by Sinorhizobium meliloti strain Rm41, BMC Plant Biol. 14 (2014) 1.

[29] S. Yang, F. Tang, M. Gao, H.B. Krishnan, H. Zhu, R gene-controlled host specificity in the legume-rhizobia symbiosis, Proc. Natl. Acad. Sci. 107 (2010) 18735–18740. doi:10.1073/pnas.1011957107.

[30] T. Kazmierczak, M. Nagymihály, F. Lamouche, Q. Barrière, I. Guefrachi, B. Alunni, M. Ouadghiri, J. Ibijbijen, É. Kondorosi, P. Mergaert, V. Gruber, Specific Host-Responsive Associations Between *Medicago truncatula* Accessions and *Sinorhizobium* Strains, Mol. Plant. Microbe Interact. 30 (2017) 399–409. doi:10.1094/MPMI-01-17-0009-R.

[31] S. Simsek, T. Ojanen-Reuhs, S.B. Stephens, B.L. Reuhs, Strain-Ecotype Specificity in *Sinorhizobium meliloti-Medicago truncatula* Symbiosis Is Correlated to Succinoglycan Oligosaccharide Structure, J. Bacteriol. 189 (2007) 7733–7740. doi:10.1128/JB.00739-07.

[32] L. Tirichine, F. de Billy, T. Huguet, Mtsym6, a Gene Conditioning *Sinorhizobium* Strain-Specific Nitrogen Fixation in *Medicago truncatula*, Plant Physiol. 123 (2000) 845–852.

[33] P. Somasegaran, H.J. Hoben, Handbook for Rhizobia: Methods in Legume-Rhizobium technology, Springer, 1994. http://www.springer.com/gp/book/9781461383772 (accessed June 12, 2017).

[34] S. Andrews, Fastqc. A quality control tool for high throughput sequence data, (2010). http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed March 2, 2016).

[35] B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, M.B. Couger, D. Eccles, B. Li, M. Lieber, M.D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C.N. Dewey, R. Henschel, R.D. LeDuc, N. Friedman, A. Regev, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis, Nat. Protoc. 8 (2013) 1494–1512. doi:10.1038/nprot.2013.084.

[36] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods. 9 (2012) 357–359. doi:10.1038/nmeth.1923.

[37] N.M. Davidson, A. Oshlack, Corset: enabling differential gene expression analysis for de novo assembled transcriptomes, Genome Biol. 15 (2014) 410.

[38] F.A. Simão, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, Bioinformatics. 31 (2015) 3210–3212. doi:10.1093/bioinformatics/btv351.

[39] R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E.L.L. Sonnhammer, J. Tate, M. Punta, Pfam: the protein families database, Nucleic Acids Res. 42 (2014) D222–D230. doi:10.1093/nar/gkt1223.

ACCEPTED MANUSCRIPT

[40] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST+: architecture and applications, BMC Bioinformatics. 10 (2009) 421. doi:10.1186/1471-2105-10-421.

[41] R.D. Finn, J. Clements, S.R. Eddy, HMMER web server: interactive sequence similarity searching, Nucleic Acids Res. 39 (2011) W29–W37. doi:10.1093/nar/gkr367.

[42] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, Nat. Methods. 8 (2011) 785–786.

[43] A. Krogh, B. Larsson, G. von Heijne, E.L.. Sonnhammer, Predicting transmembrane protein topology with a hidden markov model: application to complete genomes, J. Mol. Biol. 305 (2001) 567–580. doi:10.1006/jmbi.2000.4315.

[44] H. Tang, V. Krishnakumar, S. Bidwell, B. Rosen, A. Chan, S. Zhou, L. Gentzbittel, K.L. Childs, M. Yandell, H. Gundlach, K.F.X. Mayer, D.C. Schwartz, C.D. Town, An improved genome release (version Mt4. 0) for the model legume *Medicago truncatula*, BMC Genomics. 15 (2014) 312.

[45] J. Schmutz, S.B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D.L. Hyten, Q. Song, J.J. Thelen, J. Cheng, D. Xu, U. Hellsten, G.D. May, Y. Yu, T. Sakurai, T. Umezawa, M.K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindquist, M. Peto, D. Grant, S. Shu, D. Goodstein, K. Barry, M. Futrell-Griggs, B. Abernathy, J. Du, Z. Tian, L. Zhu, N. Gill, T. Joshi, M. Libault, A. Sethuraman, X.-C. Zhang, K. Shinozaki, H.T. Nguyen, R.A. Wing, P. Cregan, J. Specht, J. Grimwood, D. Rokhsar, G. Stacey, R.C. Shoemaker, S.A. Jackson, Genome sequence of the palaeopolyploid soybean, Nature. 463 (2010) 178–183. doi:10.1038/nature08670.

[46] D. Risso, K. Schwartz, G. Sherlock, S. Dudoit, GC-content normalization for RNA-Seq data, BMC Bioinformatics. 12 (2011) 480.

[47] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014). doi:10.1186/s13059-014-0550-8.

[48] A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, S.G. Rozen, Primer3--new capabilities and interfaces, Nucleic Acids Res. 40 (2012) e115–e115. doi:10.1093/nar/gks596.

[49] K.J. Livak, T.D. Schmittgen, Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2???CT Method, Methods. 25 (2001) 402–408. doi:10.1006/meth.2001.1262.

[50] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, F. Speleman, Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, Genome Biol. 3 (2002) research0034–1.

[51] O. Thimm, O. Bläsing, Y. Gibon, A. Nagel, S. Meyer, P. Krüger, J. Selbig, L.A. Müller, S.Y. Rhee, M. Stitt, Mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes, Plant J. 37 (2004) 914–939. doi:10.1111/j.1365-313X.2004.02016.x.

[52] B. Usadel, A. Nagel, D. Steinhauser, Y. Gibon, O.E. Bläsing, H. Redestig, N. Sreenivasulu, L. Krall, M.A. Hannah, F. Poree, A.R. Fernie, M. Stitt, PageMan: an interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments, BMC Bioinformatics. 7 (2006) 535.

[53] S. Dray, A.B. Dufour, The ade4 package: implementing the duality diagram for ecologists, J. Stat. Softw. 22 (2007) 1–20.

[54] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2017. https://www.R-project.org/.

[55] RStudio Team, RStudio: Integrated Development Environment for R, RStudio, Inc., Boston, MA, 2016. http://www.rstudio.com/.

[56] H. Wickham, ggplot2, Springer New York, New York, NY, 2009. doi:10.1007/978-0-387-98141-3.

[57] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat. Biotechnol. 29 (2011) 644–652. doi:10.1038/nbt.1883.

[58] F. Mahé, H. Pascual, O. Coriton, V. Huteau, A. Navarro Perris, M.-T. Misset, A. Aïnouche, New data and phylogenetic placement of the enigmatic Old World lupin: *Lupinus mariae-josephi* H. Pascual, Genet. Resour. Crop Evol. 58 (2011) 101–114. doi:10.1007/s10722-010-9580-6.

[59] M.-F. Jardinaud, S. Boivin, N. Rodde, O. Catrice, A. Kisiala, A. Lepage, S. Moreau, B. Roux, L. Cottret, E. Sallet, others, A laser dissection-RNAseq analysis highlights the activation of cytokinin

ACCEPTED MANUSCRIPT

pathways by nod factors in the *Medicago truncatula* root epidermis, Plant Physiol. 171 (2016) 2256–2276.

[60] S. Yuan, R. Li, S. Chen, H. Chen, C. Zhang, L. Chen, Q. Hao, Z. Shan, Z. Yang, D. Qiu, X. Zhang, X. Zhou, RNA-Seq Analysis of Differential Gene Expression Responding to Different Rhizobium Strains in Soybean (Glycine max) Roots, Front. Plant Sci. 7 (2016). doi:10.3389/fpls.2016.00721.

[61] S. Hirsch, J. Kim, A. Munoz, A.B. Heckmann, J.A. Downie, G.E.D. Oldroyd, GRAS Proteins Form a DNA Binding Complex to Induce Gene Expression during Nodulation Signaling in Medicago truncatula, Plant Cell Online. 21 (2009) 545–557. doi:10.1105/tpc.108.064501.

[62] M. Mbengue, S. Camut, F. de Carvalho-Niebel, L. Deslandes, S. Froidure, D. Klaus-Heisen, S. Moreau, S. Rivas, T. Timmers, C. Hervé, J. Cullimore, B. Lefebvre, The *Medicago truncatula* E3 Ubiquitin Ligase PUB1 Interacts with the LYK3 Symbiotic Receptor and Negatively Regulates Infection and Nodulation, Plant Cell. 22 (2010) 3474–3488. doi:10.1105/tpc.110.075861.

[63] J.D. Murray, R.R.D. Muni, I. Torres-Jerez, Y. Tang, S. Allen, M. Andriankaja, G. Li, A. Laxmi, X. Cheng, J. Wen, D. Vaughan, M. Schultze, J. Sun, M. Charpentier, G. Oldroyd, M. Tadege, P. Ratet, K.S. Mysore, R. Chen, M.K. Udvardi, Vapyrin, a gene essential for intracellular progression of arbuscular mycorrhizal symbiosis, is also essential for infection by rhizobia in the nodule symbiosis of *Medicago truncatula*, Plant J. 65 (2011) 244–252. doi:10.1111/j.1365-313X.2010.04415.x.

[64] S. Sinharoy, I. Torres-Jerez, K. Bandyopadhyay, A. Kereszt, C.I. Pislariu, J. Nakashima, V.A. Benedito, E. Kondorosi, M.K. Udvardi, The C2H2 Transcription Factor REGULATOR OF SYMBIOSOME DIFFERENTIATION Represses Transcription of the Secretory Pathway Gene VAMP721a and Promotes Symbiosome Development in *Medicago truncatula*, Plant Cell. 25 (2013) 3584–3601. doi:10.1105/tpc.113.114017.

[65] F.C. Guinel, Ethylene, a Hormone at the Center-Stage of Nodulation, Front. Plant Sci. 6 (2015). doi:10.3389/fpls.2015.01121.

[66] G. Colebatch, G. Desbrosses, T. Ott, L. Krusell, O. Montanari, S. Kloska, J. Kopka, M.K. Udvardi, Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*, Plant J. 39 (2004) 487–512. doi:10.1111/j.1365-313X.2004.02150.x.

[67] A.J. Gordon, F.R. Minchin, C.L. James, O. Komina, Sucrose synthase in legume nodules is essential for nitrogen fixation, Plant Physiol. 120 (1999) 867–878.

[68] N. Hohnjec, J.D. Becker, A. Pühler, A.M. Perlick, H. Küster, Genomic organization and expression properties of the *MtSucS1* gene, which encodes a nodule-enhanced sucrose synthase in the model legume *Medicago truncatula*, Mol. Gen. Genet. MGG. 261 (1999) 514–522.

[69] V. Krishnakumar, M. Kim, B.D. Rosen, S. Karamycheva, S.L. Bidwell, H. Tang, C.D. Town, MTGD: The *Medicago truncatula* Genome Database, Plant Cell Physiol. 56 (2015) e1–e1. doi:10.1093/pcp/pcu179.

[70] L. Brechenmacher, Z. Lei, M. Libault, S. Findley, M. Sugawara, M.J. Sadowsky, L.W. Sumner, G. Stacey, Soybean Metabolites Regulated in Root Hairs in Response to the Symbiotic Bacterium *Bradyrhizobium japonicum*, PLANT Physiol. 153 (2010) 1808–1822. doi:10.1104/pp.110.157800.

[71] A. Barraza, C. Contreras-Cubas, G. Estrada-Navarrete, J.L. Reyes, M.A. Juárez-Verdayes, N. Avonce, C. Quinto, C. Díaz-Camino, F. Sanchez, The Class II Trehalose 6-phosphate Synthase Gene PvTPS9 Modulates Trehalose Metabolism in *Phaseolus vulgaris* Nodules, Front. Plant Sci. 7 (2016). doi:10.3389/fpls.2016.01589.

[72] A. Barraza, G. Estrada-Navarrete, M.E. Rodriguez-Alegria, A. Lopez-Munguia, E. Merino, C. Quinto, F. Sanchez, Down-regulation of *PvTRE1* enhances nodule biomass and bacteroid number in the common bean, New Phytol. 197 (2013) 194–206. doi:10.1111/nph.12002.

[73] M. Betti, M. García-Calderón, C.M. Pérez-Delgado, A. Credali, G. Estivill, F. Galván, J.M. Vega, A.J. Márquez, Glutamine Synthetase in Legumes: Recent Advances in Enzyme Structure and Functional Genomics, Int. J. Mol. Sci. 13 (2012) 7994–8024. doi:10.3390/ijms13077994.

[74] P.-K. Hsu, Y.-F. Tsay, Two Phloem Nitrate Transporters, NRT1.11 and NRT1.12, Are Important for Redistributing Xylem-Borne Nitrate to Enhance Plant Growth, PLANT Physiol. 163 (2013) 844–856. doi:10.1104/pp.113.226563.

[75] M. Tejada-Jiménez, R. Castro-Rodríguez, I. Kryvoruchko, M.M. Lucas, M. Udvardi, J. Imperial, M. González-Guerrero, *Medicago truncatula* Natural Resistance-Associated Macrophage Protein1 Is Required for Iron Uptake by Rhizobia-Infected Nodule Cells, Plant Physiol. 168 (2015) 258–272. doi:10.1104/pp.114.254672.

[76] C. Tang, A.D. Robson, M.J. Dilworth, The role of iron in nodulation and nitrogen fixation in *Lupinus angustifolius* L, New Phytol. 114 (1990) 173–182. doi:10.1111/j.1469-8137.1990.tb00388.x.

ACCEPTED MANUSCRIPT

[77] G.E.D. Oldroyd, J.D. Murray, P.S. Poole, J.A. Downie, The Rules of Engagement in the Legume-Rhizobial Symbiosis, Annu. Rev. Genet. 45 (2011) 119–144. doi:10.1146/annurev-genet-110410-132549.

[78] E. Larrainzar, B.K. Riely, S.C. Kim, N. Carrasquilla-Garcia, H.-J. Yu, H.-J. Hwang, M. Oh, G.B. Kim, A.K. Surendrarao, D. Chasman, A.F. Siahpirani, R.V. Penmetsa, G.-S. Lee, N. Kim, S. Roy, J.-H. Mun, D.R. Cook, Deep Sequencing of the *Medicago truncatula* Root Transcriptome Reveals a Massive and Early Interaction between Nodulation Factor and Ethylene Signals, Plant Physiol. 169 (2015) 233–265. doi:10.1104/pp.15.00350.

[79] P.H. Middleton, J. Jakab, R.V. Penmetsa, C.G. Starker, J. Doll, P. Kalo, R. Prabhu, J.F. Marsh, R.M. Mitra, A. Kereszt, B. Dudas, K. VandenBosch, S.R. Long, D.R. Cook, G.B. Kiss, G.E.D. Oldroyd, An ERF Transcription Factor in *Medicago truncatula* That Is Essential for Nod Factor Signal Transduction, Plant Cell. 19 (2007) 1221–1234. doi:10.1105/tpc.106.048264.

[80] W. Grunewald, G. van Noorden, G. Van Isterdael, T. Beeckman, G. Gheysen, U. Mathesius, Manipulation of Auxin Transport in Plant Roots during Rhizobium Symbiosis and Nematode Parasitism, Plant Cell Online. 21 (2009) 2553–2562. doi:10.1105/tpc.109.069617.

[81] E. Di Giacomo, C. Laffont, F. Sciarra, M.A. Iannelli, F. Frugier, G. Frugis, KNAT3/4/5-like class 2 KNOX transcription factors are involved in *Medicago truncatula* symbiotic nodule organ development, New Phytol. 213 (2017) 822–837. doi:10.1111/nph.14146.

[82] V. Mortier, A. Wasson, P. Jaworek, A. De Keyser, M. Decroos, M. Holsters, P. Tarkowski, U. Mathesius, S. Goormachtig, Role of *LONELY GUY* genes in indeterminate nodulation on *Medicago truncatula*, New Phytol. 202 (2014) 582–593. doi:10.1111/nph.12681.

[83] T. Werner, T. Schmülling, Cytokinin action in plant development, Curr. Opin. Plant Biol. 12 (2009) 527–538. doi:10.1016/j.pbi.2009.07.002.

[84] T. Vernié, S. Moreau, F. de Billy, J. Plet, J.-P. Combier, C. Rogers, G. Oldroyd, F. Frugier, A. Niebel, P. Gamas, EFD Is an ERF Transcription Factor Involved in the Control of Nodule Number and Differentiation in *Medicago truncatula*, Plant Cell Online. 20 (2008) 2696–2713. doi:10.1105/tpc.108.059857.

[85] S. Boivin, T. Kazmierczak, M. Brault, J. Wen, P. Gamas, K.S. Mysore, F. Frugier, Different cytokinin histidine kinase receptors regulate nodule initiation as well as later nodule developmental stages in *Medicago truncatula*: Specificity and redundancy of CHKs in nodulation, Plant Cell Environ. 39 (2016) 2198–2209. doi:10.1111/pce.12779.

[86] Y. Ding, P. Kalo, C. Yendrek, J. Sun, Y. Liang, J.F. Marsh, J.M. Harris, G.E.D. Oldroyd, Abscisic Acid Coordinates Nod Factor and Cytokinin Signaling during the Regulation of Nodulation in *Medicago truncatula*, Plant Cell Online. 20 (2008) 2681–2695. doi:10.1105/tpc.108.061739.

[87] S. Fuchs, S.V. Tischer, C. Wunschel, A. Christmann, E. Grill, Abscisic acid sensor RCAR7/PYL13, specific regulator of protein phosphatase coreceptors, Proc. Natl. Acad. Sci. 111 (2014) 5741–5746. doi:10.1073/pnas.1322085111.

[88] H. -i. Choi, Arabidopsis Calcium-Dependent Protein Kinase AtCPK32 Interacts with ABF4, a Transcriptional Regulator of Abscisic Acid-Responsive Gene Expression, and Modulates Its Activity, PLANT Physiol. 139 (2005) 1750–1761. doi:10.1104/pp.105.069757.

[89] B.J. Ferguson, E. Foo, J.J. Ross, J.B. Reid, Relationship between gibberellin, ethylene and nodulation in Pisum sativum, New Phytol. 189 (2011) 829–842. doi:10.1111/j.1469-8137.2010.03542.x.

[90] S. Hayashi, P.M. Gresshoff, B.J. Ferguson, Mechanistic action of gibberellins in legume nodulation: Gibberellin and nodulation, J. Integr. Plant Biol. 56 (2014) 971–978. doi:10.1111/jipb.12201.

[91] M. Libault, A. Farmer, L. Brechenmacher, J. Drnevich, R.J. Langley, D.D. Bilgin, O. Radwan, D.J. Neece, S.J. Clough, G.D. May, G. Stacey, Complete Transcriptome of the Soybean Root Hair Cell, a Single-Cell Model, and Its Alteration in Response to *Bradyrhizobium japonicum* Infection, PLANT Physiol. 152 (2010) 541–552. doi:10.1104/pp.109.148379.

[92] Y. Jin, H. Liu, D. Luo, N. Yu, W. Dong, C. Wang, X. Zhang, H. Dai, J. Yang, E. Wang, DELLA proteins are common components of symbiotic rhizobial and mycorrhizal signalling pathways, Nat. Commun. 7 (2016) 12433. doi:10.1038/ncomms12433.

[93] J. Park, K.T. Nguyen, E. Park, J.-S. Jeon, G. Choi, DELLA Proteins and Their Interacting RING Finger Proteins Repress Gibberellin Responses by Binding to the Promoters of a Subset of Gibberellin-Responsive Genes in *Arabidopsis*, Plant Cell. 25 (2013) 927–943. doi:10.1105/tpc.112.108951.

[94] M.L. Falcone Ferreyra, S.P. Rius, P. Casati, Flavonoids: biosynthesis, biological functions, and biotechnological applications, Front. Plant Sci. 3 (2012). doi:10.3389/fpls.2012.00222.

ACCEPTED MANUSCRIPT

[95] A.P. Wasson, Silencing the Flavonoid Pathway in *Medicago truncatula* Inhibits Root Nodule Formation and Prevents Auxin Transport Regulation by Rhizobia, Plant Cell Online. 18 (2006) 1617–1629. doi:10.1105/tpc.105.038232.

[96] J. Zhang, S. Subramanian, G. Stacey, O. Yu, Flavones and flavonols play distinct critical roles during nodulation of *Medicago truncatula* by *Sinorhizobium meliloti*, Plant J. 57 (2009) 171–183. doi:10.1111/j.1365-313X.2008.03676.x.

[97] S. Subramanian, G. Stacey, O. Yu, Distinct, crucial roles of flavonoids during legume nodulation, Trends Plant Sci. 12 (2007) 282–285.

[98] H. Gagnon, J. Grandmaison, R.K. Ibrahim, Phytochemical and immunocytochemical evidence for the accumulation of 2'-hydroxylupalbigenin in lupin nodules and bacteroids., Mol. Plant Microbe Interact. MPMI. 8 (1995) 131–137.

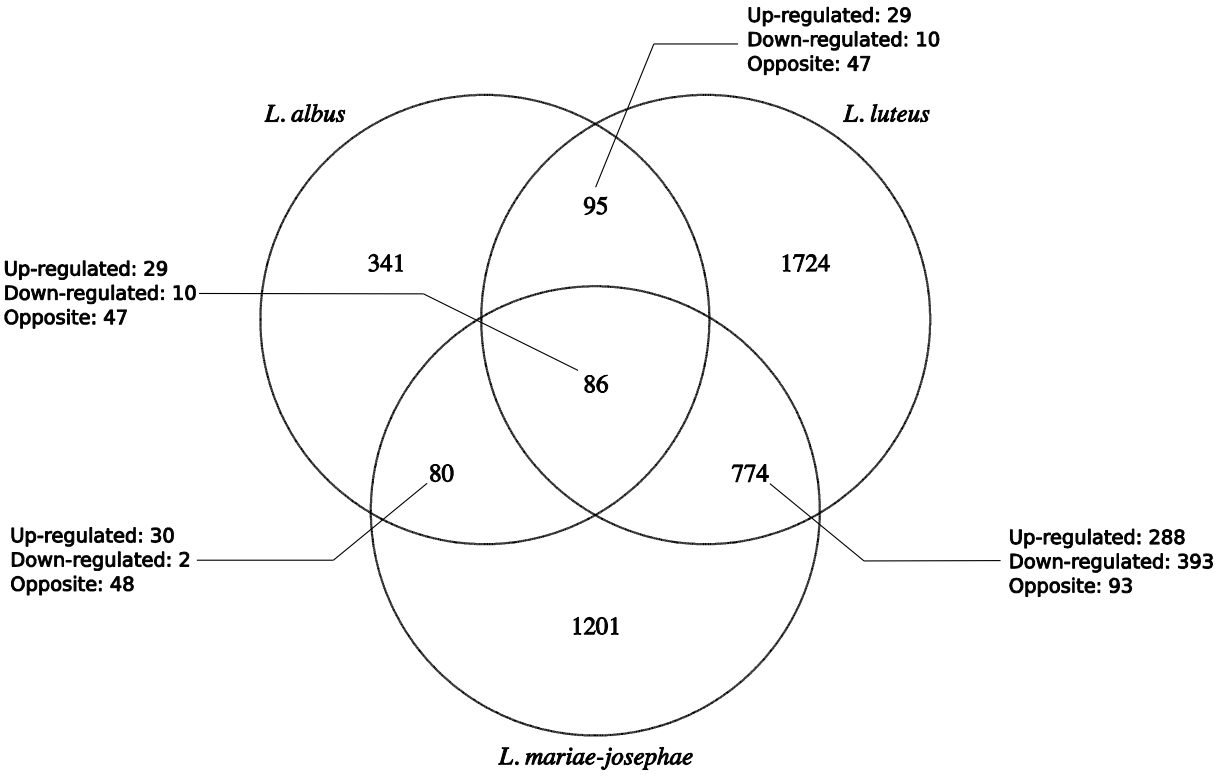
[99] J. Grandmaison, R. Ibrahim, Ultrastructural localization of a diprenylated isoflavone in Rhizobium lupini-Lupinus albus symbiotic association, J. Exp. Bot. 46 (1995) 231–237.

[100] E. Miedes, R. Vanholme, W. Boerjan, A. Molina, The role of the secondary cell wall in plant resistance to pathogens, Front. Plant Sci. 5 (2014). doi:10.3389/fpls.2014.00358.

ACCEPTED MANUSCRIPT

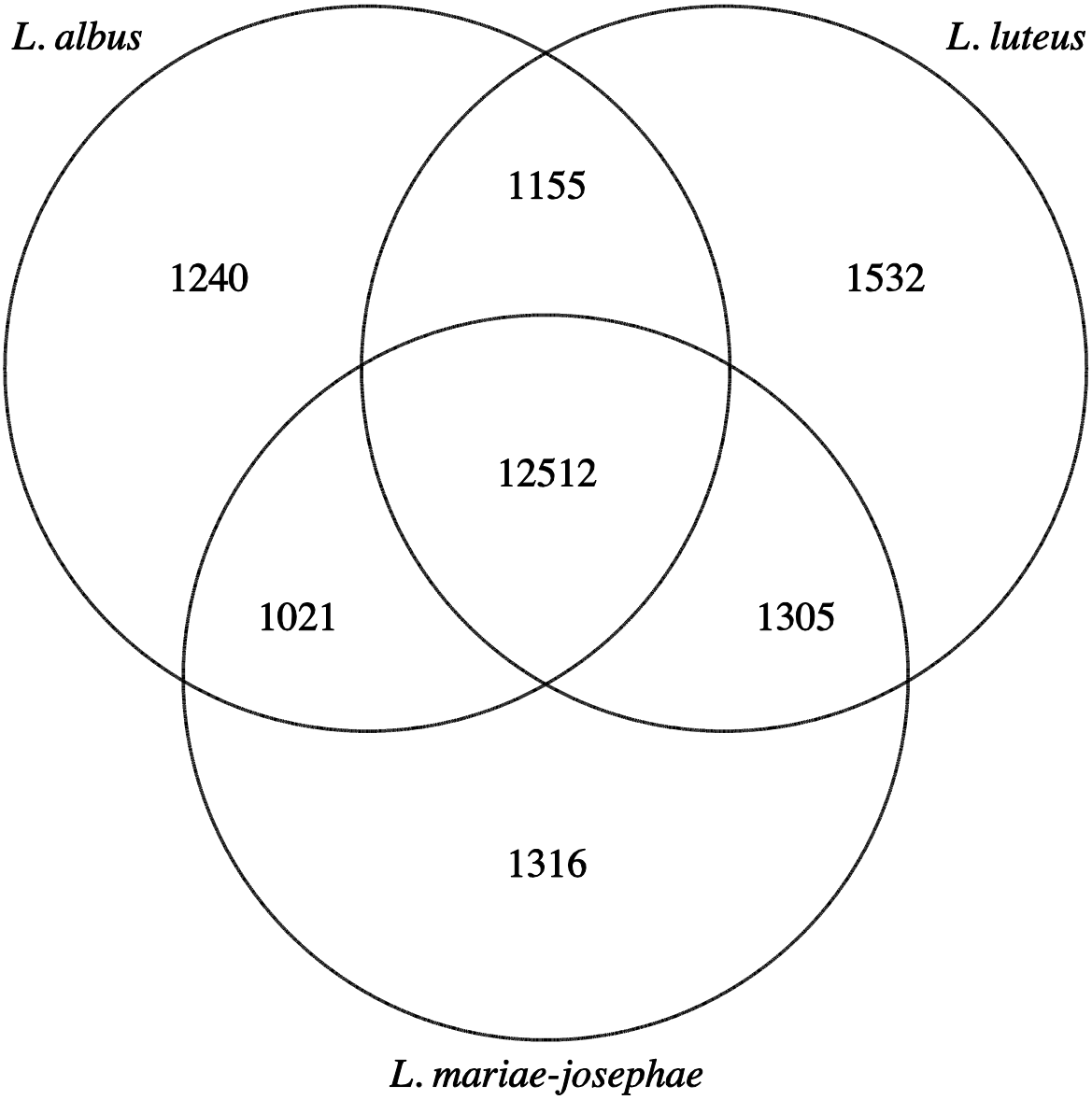
Figure legends

Figure 1: Venn diagram of shared and specific assigned DE clusters across three lupine species, among a total of 602 clusters in *L. albus*, 2,679 in *L. luteus* and 2,141 in *L. mariae-josephae*. The expression direction is indicated for shared clusters (opposite means that a cluster did not have the same expression direction in the considered species).



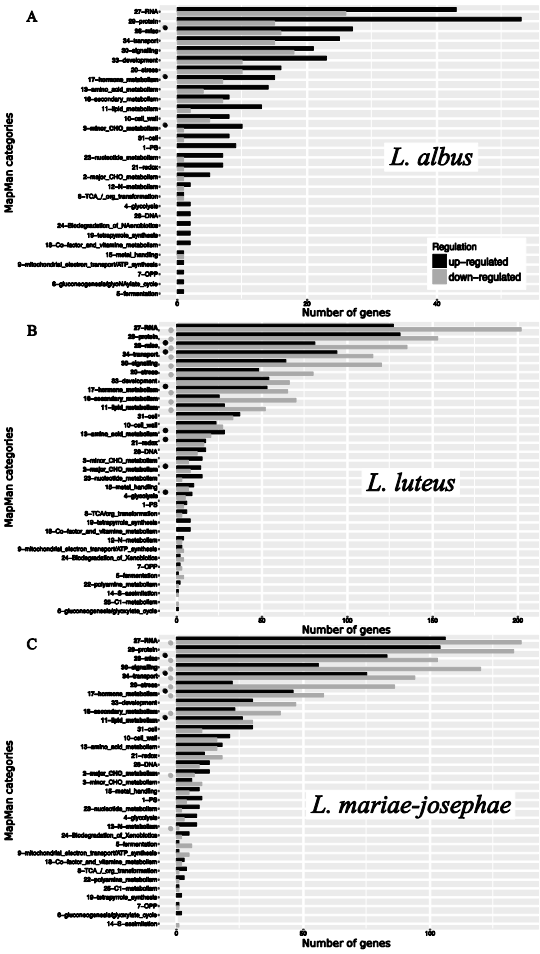
ACCEPTED MANUSCRIPT

Figure 2: Venn diagram of shared and specific clusters assigned to a unique ID of *M. truncatula* across the three lupine species. Total of 15,928 clusters in *L. albus*, 16,504 in *L. luteus* and 16,154 in *L. mariae-josephae*.



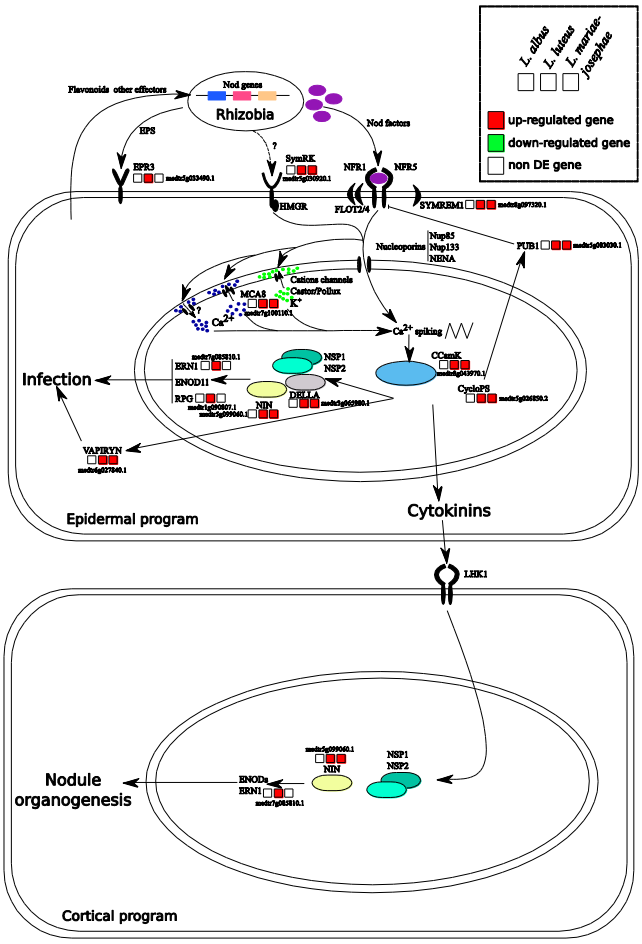
ACCEPTED MANUSCRIPT

Figure 3: Functional classification of the differentially expressed clusters in A. *L. albus*; B. *L. luteus* and C. *L. mariae-josephae*, between compatible *vs* incompatible symbiotic interactions. Up-regulated genes are indicated by black bars, down-regulated genes by grey bars. A black or grey circle at the base of the graph indicates the over-representation of up- or down-regulated genes, respectively. Over-represented functional categories are also marked in bold.



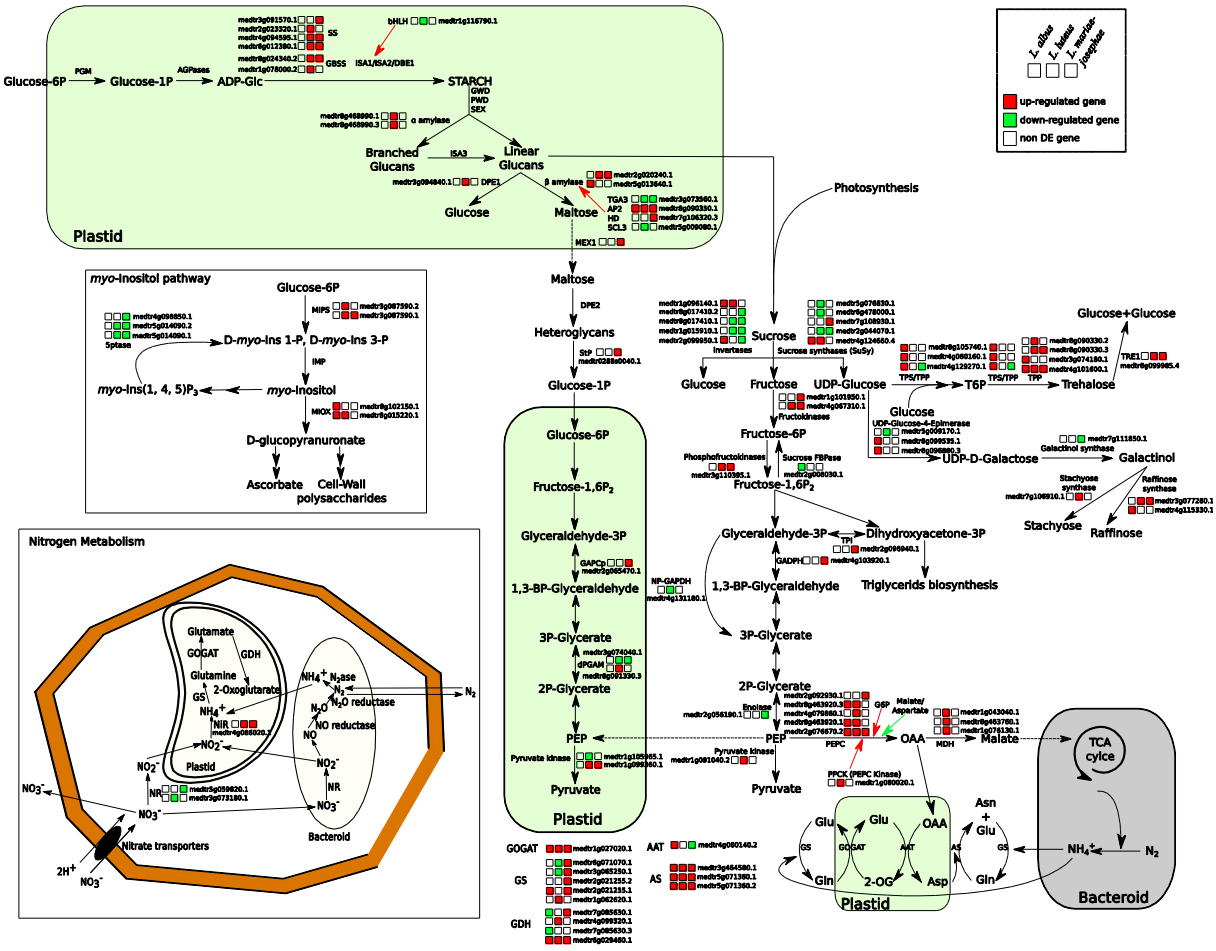
ACCEPTED MANUSCRIPT

Figure 4: Simplified regulatory pathway of plant-bacteria recognition, epidermal and cortical programs. For each gene, lupine species *L. albus*, *L. luteus* and *L. mariae-josephae* are respectively represented by the first, second and third squares. Squares are filled according the gene expression pattern as indicated in the top-right dashed rectangle: dark filling (red) stands for up-regulated genes (induced by compatible strains), light filling (green) for down-regulated genes (induced by incompatible strains) and white for genes not differentially expressed between the compatible and the non-compatible symbiotic situations. For each gene, the *M. truncatula* ortholog ID is indicated. Figure adapted from [7,8].



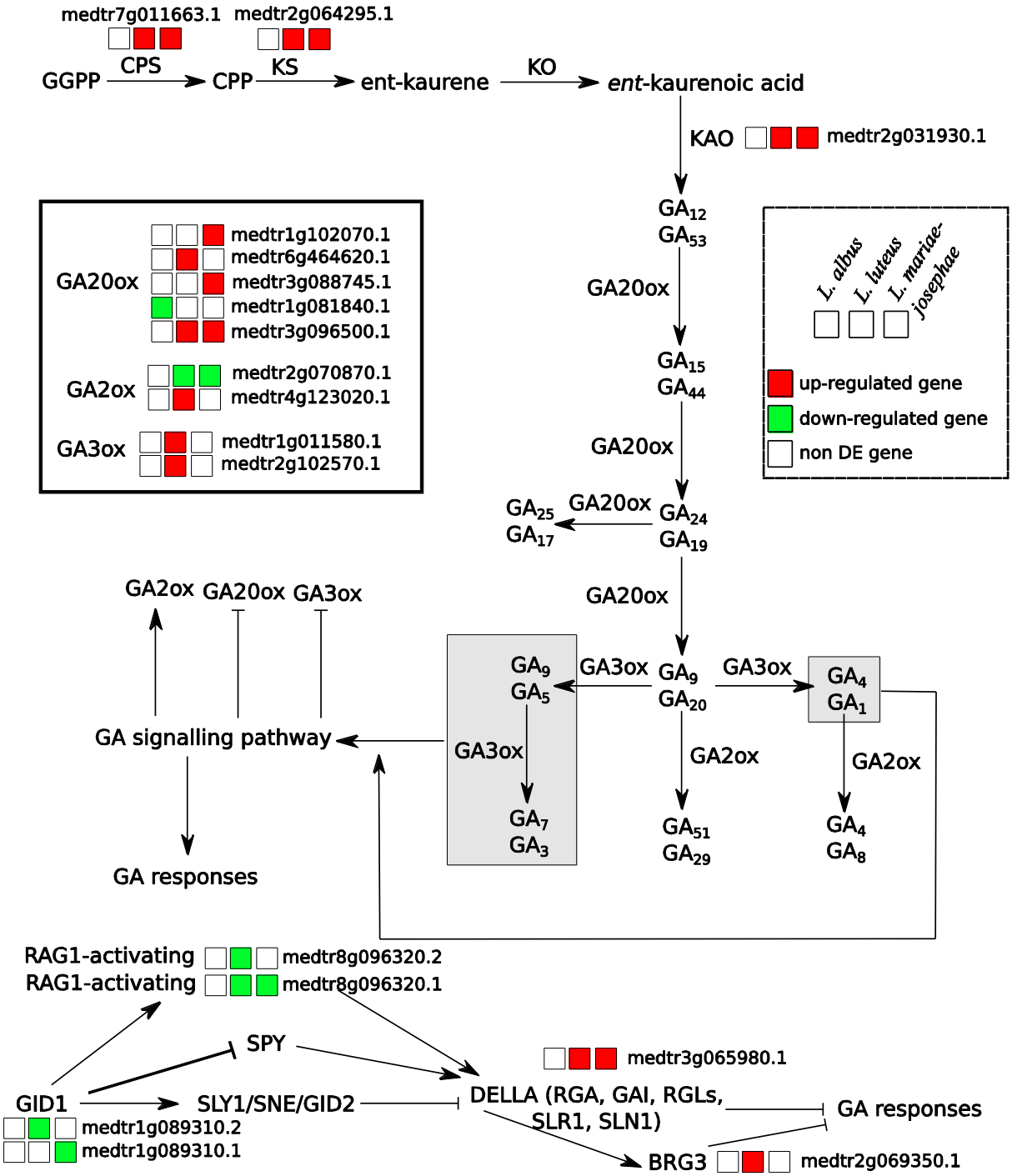
ACCEPTED MANUSCRIPT

Figure 5: Carbon and nitrogen metabolisms and nutrient exchanges between plant and atmospheric nitrogen-fixing bacteroid. For each gene, lupine species *L. albus*, *L. luteus* and *L. mariae-josephae* are respectively represented by the first, second and third squares. Squares are filled according the gene expression pattern as indicated in the top-right dashed rectangle: dark filling (red) stands for up-regulated genes (induced by compatible strains), light filling (green) for down-regulated genes (induced by incompatible strains) and white for not differentially expressed genes between the compatible and the non-compatible symbiotic situations. For each gene, the *M. truncatula* ortholog ID is indicated. Figure adapted from [73].



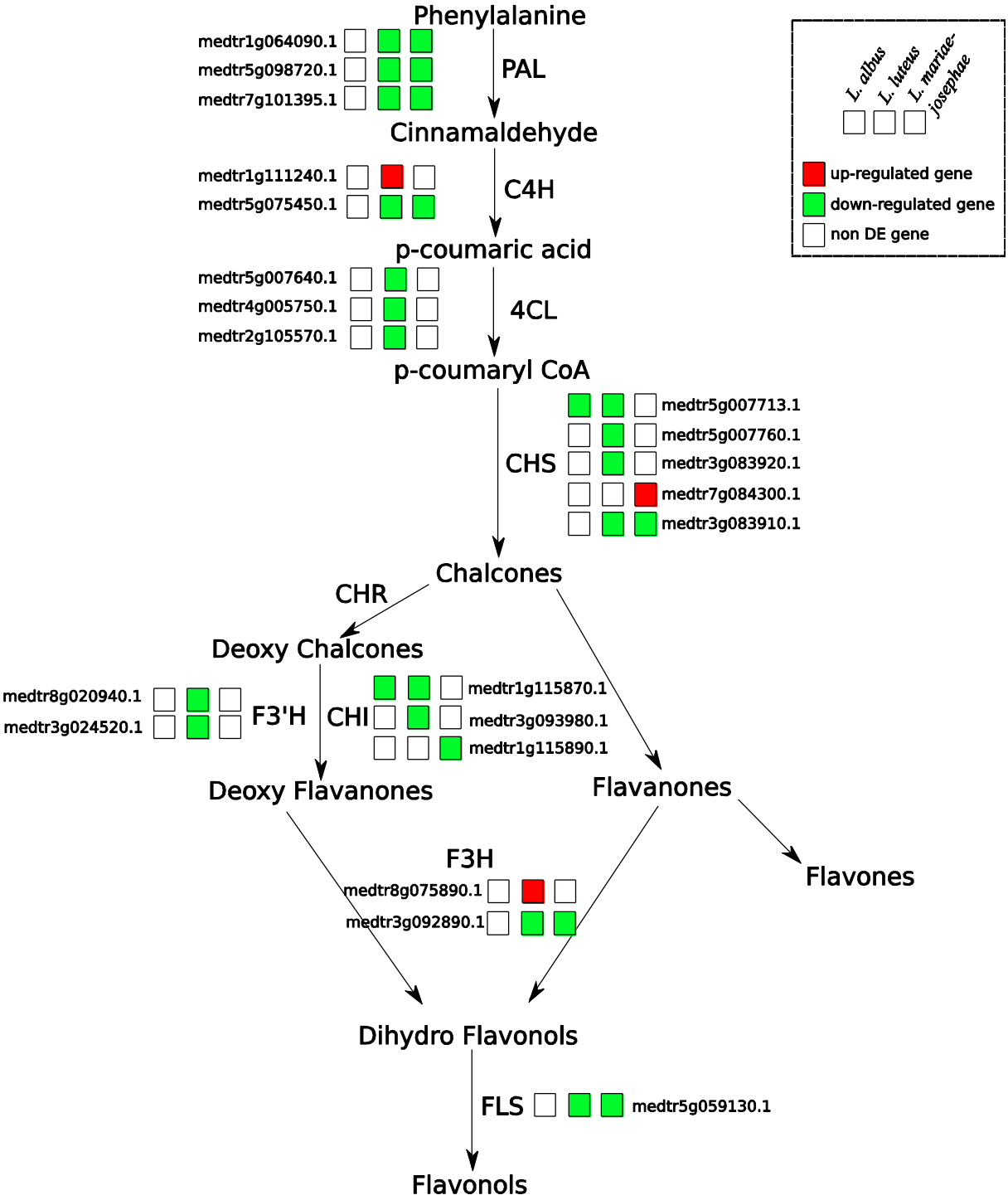
ACCEPTED MANUSCRIPT

Figure 6: Gibberellins biosynthesis pathway and downstream regulation of GAs responses. For each gene, lupine species *L. albus*, *L. luteus* and *L. mariae-josephae* are respectively represented by the first, second and third squares. Squares are filled according the gene expression pattern as indicated in the top-right dashed rectangle: dark filling (red) stands for up-regulated genes (induced by compatible strains), light filling (green) for down-regulated genes (induced by incompatible strains) and white for not differentially expressed genes between the compatible and the non-compatible symbiotic situations. For each gene, the *M. truncatula* ortholog ID is indicated. Grey boxes indicate the biologically active forms of GAs. Expression pattern of genes encoding GA20ox, GA3ox and GA2ox are indicated in the dashed square. Figure adapted from [89,90].



ACCEPTED MANUSCRIPT

Figure 7: Flavonoids biosynthesis pathway. For each gene, lupine species *L. albus*, *L. luteus* and *L. mariae-josephae* are respectively represented by the first, second and third squares. Squares are filled according the gene expression pattern as indicated in the top-right dashed rectangle: dark filling (red) stands for up-regulated genes (induced by compatible strains), light filling (green) for down-regulated genes (induced by incompatible strains) and white for not differentially expressed genes between the compatible and the non-compatible symbiotic situations. For each gene, the *M. truncatula* ortholog ID is indicated. Only flavonoids biosynthetic pathways containing DE genes are represented here. Figure adapted from [94].



ACCEPTED MANUSCRIPT

Table 1: Description of Illumina sequencing, assembly, clustering and expression analysis results. PE: Paired-End; DE: Differentially Expressed.

	L. albus	*L. luteus*	*L. mariae-josephae*
Total number of 150bp PE reads	214,793,333	224,368,938	231,015,211
Number of Trinity contigs	170,622	192,835	184,265
Number of clusters	119,403	132,380	130,322
Number of clusters after expression filter	37,935	42,248	38,320
Number of DE clusters (p value <0.05)	1,206	4,790	4,582
Number of DE clusters (p value <0.05 and 1<log2FC<-1)	1,146	4,759	3,978